

GENES

**BENJAMIN
• LEWIN •**

**EDITION
SPECIALE
ETUDIANT**

Médecine-Sciences
Flammarion

BIBL. CENTR. UNIV.
M. „FELIX RUPESCU”
BUCUREȘTI
IV. 13.019

GENES



100000
BCU 1000

10000000000000000000
10000000000000000000

BCU IASI / CENTRAL UNIVERSITY LIBRARY

GENES

BENJAMIN LEWIN

Traduit de l'américain par **AXEL KAHN**



149812
B.C.U. - IASI

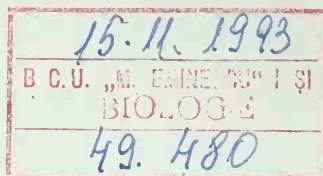
Médecine-Sciences
Flammarion

Ont participé à cette traduction, sous la direction de Axel Kahn :

Pascale Briand, Danièle Choury,
Daniel Corcos, Nathalie Denis, Nathalie Dubois,
Gisèle Grimber, Jean-Jacques Pernelle,
Daniel Rabier, Marie-France Szajnert.

1^{re} édition, 1988

2^e tirage, 1992



Cet ouvrage est paru dans son édition américaine
sous le titre : Genes III.

Illustrations by John Balbalis with the assistance of the
Wiley Illustration Department.

Copyright © 1983, 1985, 1987 by John Wiley & Sons, Inc.

All rights reserved.

Authorized translation from English language edition published
by John Wiley & Sons, inc.

ISBN : 2-257-15041-4

© 1988, 1992 by Flammarion

Printed in France

Pour recevoir le catalogue Flammarion Médecine-Sciences,
il suffit d'envoyer vos nom et adresse à :

Flammarion Médecine-Sciences

4, rue Casimir-Delavigne,
75006 Paris.

PRÉFACE

Résumé des connaissances actuelles en biologie moléculaire, le but de *GÈNES* est de proposer une vision cohérente de la génétique, partant de ce que l'on sait du gène. A la lumière des découvertes et des techniques les plus récentes, *GÈNES* est le reflet d'une conception nouvelle de la génétique considérée comme un tout. Selon moi, étudier la génétique signifie principalement, à l'heure actuelle, que l'on travaille sur l'ADN.

Le thème sous-jacent à *GÈNES*, repris édition par édition, est que le gène a à sa disposition un vaste répertoire de stratégies permettant la survie des espèces ; différents exemples en seront présentés dans des systèmes variés. A l'extrême, on pourrait dire que l'organisme (n'importe quel organisme) est avant tout une trousse de survie permettant la conservation des gènes qu'il porte.

Cet ouvrage, comme les éditions précédentes, souligne ce qui revient à un nouveau domaine des connaissances, dont l'extraordinaire progression a pratiquement écrasé la génétique traditionnelle. En réunissant en un format acceptable l'énorme masse d'informations qui s'est accumulée, il est maintenant possible de poser les questions fondamentales : qu'est-ce qu'un gène, comment se reproduit-il, comment s'exprime-t-il et qu'est-ce qui contrôle son expression ?

GÈNES traite d'abord de l'expression d'un gène en une protéine et, de la protéine, nous travaillons au niveau moléculaire, pour ainsi dire à reculons, vers l'ADN lui-même. Pour refléter les perspectives de la recherche actuelle, nous donnons une égale importance à la biologie moléculaire des procaryotes et des eucaryotes. L'une et l'autre sont maintenant parties intégrantes de la même histoire. Dans certains cas, il semble légitime de traiter ensemble des deux systèmes, alors que dans d'autres cas, des chapitres séparés sont consacrés aux systèmes procaryotiques et eucaryotiques.

La publication maintenant d'une nouvelle édition reflète la vitesse à laquelle sont réalisés les progrès dans ce domaine. J'ai pu constater à la lecture des articles de recherche soumis à *Cell*, ou publiés dans d'autres journaux depuis la parution de la seconde édition de *GÈNES*, que la biologie moléculaire continue

à avancer d'un pas rapide. Ce n'est cependant que pendant l'écriture de cette troisième édition que l'accélération des progrès m'est apparue si évidente. Les observations de significations essentielles continuent de s'accumuler, et ce n'est qu'en tentant de les réunir que l'on prend conscience de toute la puissance de cette progression. De ce fait, quoique *GÈNES II* ne date que de deux ans, *GÈNES III* contient une quantité considérable de nouvelles informations et propose de significatives modifications des concepts.

Les modifications dans l'organisation de cet ouvrage, introduites depuis la première édition, témoignent de mon opinion que le domaine scientifique, qui est discuté dans ce volume, est devenu l'objet principal de la recherche en biologie moderne. Afin de rendre ce *tour d'horizon* plus intelligible à un auditoire étendu, les notions de base ont été développées et de nouvelles explications ont été ajoutées au texte, là où cela semblait nécessaire.

GÈNES commence en considérant les bases biochimiques de l'hérédité, telles qu'elles apparaissent à travers la structure de l'ADN. A la suite de ce rappel des informations de base, il devrait être possible aux lecteurs dépourvus de larges connaissances de biochimie ou de génétique d'utiliser cet ouvrage un peu comme un livre ordinaire. Cette approche intègre le fait que ce livre est de plus en plus traité comme une introduction à la génétique. En effet, la plus grande partie des ouvrages, constituant une introduction à la génétique, se concentrent habituellement maintenant sur l'analyse du génome, ce qui est précisément le champ couvert par *GÈNES*.

Dans le corps principal de l'ouvrage, les discussions à propos de la transcription et de sa régulation — dont il apparaît de plus en plus évident encore qu'auparavant, qu'elle est en effet la cible principale du contrôle de l'expression des gènes — ont été réunies dans une même section. On pourrait probablement dire plus que je ne le fais sur la maturation de l'ARN chez les eucaryotes. Nous avons maintenant une vision plus précise des mouvements d'ADN dans le matériel génétique. Les méthodes de manipulation de l'ADN dans le génome sont maintenant plus puissantes. Enfin, une nouvelle et dernière section est consacrée aux discussions en cours sur le développement normal durant l'embryogenèse et sur le développement anormal des cellules cancéreuses.

Mis à part un important effort pour se concentrer sur les sujets principaux, en éliminant ou en abrégeant les domaines plus accessoires, *GÈNES III* conserve la même approche et le même style que dans les éditions antérieures. Tous les points importants sont illustrés par des diagrammes, et, quand cela est justifié, les illustrations s'efforcent de donner une idée de l'échelle et des inter-relations des éléments impliqués. Comme auparavant, chaque chapitre se termine par une brève revue bibliographique qui propose certaines synthèses récentes utiles et certains articles de recherche particulièrement significatifs permettant d'approfondir le sujet.

C'est pour moi, à nouveau, un grand plaisir de remercier mes collègues qui ont accepté de revoir certains chapitres : Welcome Bender, Michael Bishop, Arg Efstratiadis, Howard Green, Ira Herskowitz, Paula Kiberstis, Nancy Kleckner, Marilyn Kozak, James Lake, David Livingston, Ken Marcu, Howard Nash et Harold Varmus m'ont suggéré de nombreuses améliorations à apporter à la présente édition. Enfin, c'est naturellement un plaisir pour moi de remercier John Balbalis pour son aide ; ses qualités artistiques sont à l'origine de la clarification du texte par les illustrations.

BENJAMIN LEWIN

PRÉFACE À L'ÉDITION FRANÇAISE

Il est, dans l'histoire des sciences, des périodes privilégiées durant lesquelles le rythme d'acquisition des connaissances est incommensurablement plus élevé que dans les périodes précédentes. L'astronomie, la mécanique, l'optique, la physique des particules, l'électronique, l'informatique connurent ainsi de tels phénomènes d'accélération des découvertes, le plus souvent déclenchés par la mise au point d'un outil — conceptuel ou matériel — permettant de poser des questions et d'entreprendre des travaux hors de portée auparavant.

Depuis le milieu des années 1970, c'est à un tel mouvement, d'une exceptionnelle amplitude, que l'on assiste en biologie, là encore grâce à la création d'une série d'outils réunis dans ce que l'on appelle le génie génétique ou, plus proprement, la technologie des recombinants d'ADN. Ces outils ont, au départ, été créés par les « gardiens des temples », scientifiques persévérant dans l'analyse de systèmes que la mode de l'époque tendait à délaisser : la génétique et la biochimie bactérienne d'abord, puis la génétique de la drosophile et de la levure. En retour, ces outils — les enzymes de restriction, le clonage moléculaire et la détermination de la séquence de l'ADN — ont permis de faire des percées inimaginables dans la jungle du matériel génétique, celui des bactéries, des eucaryotes unicellulaires, des animaux et des plantes.

Puis, la technologie des recombinants d'ADN a diffusé au-delà des laboratoires de biologie fondamentale pour venir féconder l'ensemble des disciplines appartenant aux sciences de la vie : médecine, cancérologie, neurobiologie, immunologie, physiologie, etc.

En agriculture, la possibilité de modifier les espèces par recombinaison génétique dirigée ouvre des perspectives considérables dans le monde

végétal aussi bien qu'animal. Appliquée à l'homme, cette notion « d'amélioration des espèces » fait naître de légitimes appréhensions quant à la capacité qu'aura l'homme de maîtriser son pouvoir et son savoir. La perspective de traiter des maladies génétiques constamment incurables de nos jours représente le versant positif de ces nouvelles possibilités. Si on ajoute à l'énumération qui précède l'utilisation du génie génétique pour créer de nouveaux médicaments et vaccins... et pour dépister par leur « empreinte génétique » les criminels, on sera convaincu que la biologie moderne est devenue un fait majeur de société.

Un livre comme « GÈNES », écrit par un scientifique de talent qui est aussi le rédacteur en chef de la plus importante revue de génétique moléculaire au monde (Cell) et se trouve ainsi en première ligne pour suivre la progression de ce domaine, constitue donc une clé permettant au lecteur de s'introduire dans la fusée à étages multiples de la biologie moderne et d'être « du voyage ». J'ai cherché, en coordonnant la traduction de cet ouvrage, à restituer, sous une forme intelligible pour le lecteur français, l'esprit et les « partis pris » de l'auteur. Là où cela m'a semblé indispensable, j'ai actualisé l'information apportée par l'ajout de notes du traducteur (N. d. T.). Il me semble qu'ainsi GÈNES constitue, en français, l'ouvrage le plus complet et le plus approfondi sur les fondements et les tendances de la génétique moléculaire moderne existant à l'heure actuelle. Il illustre magnifiquement l'unité et la diversité du monde vivant ; il pénètre profondément non seulement dans l'intimité des processus génétiques complexes, mais aussi parfois dans le cheminement intellectuel et expérimental qui a conduit à leur élucidation.

En ce sens, GÈNES est une parfaite introduction à la biologie future et à ses innombrables conséquences sur les sciences et la société, pour les acteurs de cette recherche aussi bien que pour ses observateurs.

AXEL KAHN

SOMMAIRE ABRÉGÉ

INTRODUCTION

LES CELLULES OBEÏSSENT AUX LOIS DE LA PHYSIQUE ET DE LA CHIMIE 3

PARTIE 1

L'ADN, UNE RÉSERVE D'INFORMATION

15

1 LES GÈNES SONT DES UNITÉS MUTABLES

17

2 L'ADN EST LE MATÉRIEL GÉNÉTIQUE

37

3 LA TOPOLOGIE DES ACIDES NUCLÉIQUES

57

4 ISOLEMENT DU GÈNE

72

PARTIE 2

TRANSFORMER LES GÈNES EN PROTÉINES

95

5 LA CHAÎNE DE MONTAGE DE LA SYNTHÈSE PROTÉIQUE

97

6 L'ARN DE TRANSFERT : L'ADAPTATEUR TRADUCTIONNEL

122

7 LE RIBOSOME : USINE DE LA TRADUCTION

144

8 L'ARN MESSAGER MATRICE

159

PARTIE 3**CONTRÔLE DE L'EXPRESSION DES GÈNES
PAR LA TRANSCRIPTION**

	181
9 LES INTERACTIONS ARN POLYMÉRASE-PROMOTEUR CONTRÔLENT L'INITIATION	183
10 UNE PANOPLIE D'OPÉONS : LE MODÈLE DE L'OPÉON LACTOSE ET D'AUTRES EXEMPLES	219
11 CONTRÔLE AU NIVEAU DE LA TERMINAISON : L'ATTÉNUATION ET L'ANTITERMINAISON	248
12 LES CASCADES LYTQUES ET LA RÉPRESSION LYSOGÉNIQUE	269

PARTIE 4**LA PERPÉTUATION DE L'ADN**

	293
13 LE RÉPLICON, UNITÉ DE RÉPLICATION	295
14 L'APPAREIL DE RÉPLICATION DE L'ADN	312
15 LES MÉCANISMES DE MAINTENANCE DE L'ADN	335

PARTIE 5**CONSTITUTION DU GÉNOME DES EUCARYOTES**

	351
16 L'EXTRAORDINAIRE PUISSANCE DE LA TECHNOLOGIE DES RECOMBINANTS D'ADN	353
17 UN CONTINUUM DE SÉQUENCES RENFERME LES GÈNES DE STRUCTURE	367
18 L'ORGANISATION DES GÈNES MORCELÉS	384

PARTIE 6**DES GROUPES DE SÉQUENCES APPARENTÉES**

	397
19 LES GÈNES DE STRUCTURE APPARTIENNENT À DES FAMILLES DE TAILLES DIFFÉRENTES	399
20 DES GÉNOMES CONFINÉS DANS LES ORGANITES	422
21 ORGANISATION DE L'ADN DE SÉQUENCE SIMPLE	432

PARTIE 7**VERS LE PRODUIT FINAL : LA MATURATION DES ARN** 443

22	EXCISION ET RABOTAGE D'UN ARN STABLE	445
23	L'ARN EST UN CATALYSEUR : LES MÉCANISMES DE L'ÉPISSAGE	458
24	LE CONTRÔLE DE LA MATURATION DE L'ARN	483

PARTIE 8**LA CONDENSATION DE L'ADN** 497

25	AU SUJET DES GÉNOMES ET DES CHROMOSOMES	499
26	STRUCTURE DE LA CHROMATINE : LE NUCLÉOSOME	519
27	LA NATURE DE LA CHROMATINE ACTIVE	538

PARTIE 9**LE GÉNOME DYNAMIQUE : L'ADN EN MOUVEMENT** 561

28	RECOMBINAISON ET AUTRES MANIPULATIONS TOPOLOGIQUES	563
29	LES ÉLÉMENTS TRANSPOSABLES CHEZ LES BACTÉRIES	589
30	LES ÉLÉMENTS MOBILES CHEZ LES EUCARYOTES	605
31	MODIFICATIONS DE L'ORGANISATION DU GÉNOME	625

PARTIE 10**LES GÈNES AU COURS DU DÉVELOPPEMENT** 639

32	RÉARRANGEMENTS ET CONSTITUTION DE LA DIVERSITÉ IMMUNITAIRE	641
33	CHANGER L'ORGANISATION GÉNÉTIQUE DU DEDANS ET DU DEHORS	664
34	LA RÉGULATION DES GÈNES : LE CHANGEMENT DES PROGRAMMES D'EXPRESSION	681
35	ONCOGÈNES : EXPRESSION ANORMALE DES GÈNES ET CANCER	698

EPILOGUE

DÉCOUVERTES MAJEURES QUI ONT CHANGÉ NOTRE VISION DE LA GÉNÉTIQUE	719
---	------------

SOMMAIRE

INTRODUCTION

LES CELLULES OBÉISSENT AUX LOIS DE LA PHYSIQUE ET DE LA CHIMIE	3
L'HÉRÉDITÉ FONCTIONNE PAR L'ENTREMISE DE MACROMOLÉCULES	3
LES PROTÉINES SONT DES ENCHAÎNEMENTS D'ACIDES AMINÉS	4
LES FORCES NON COVALENTES SONT IMPORTANTES DANS LA CONFORMATION DES PROTÉINES	7
LES STRUCTURES PROTÉIQUES SONT EXTRÊMEMENT DIVERSES	9
COMMENT LES PROTÉINES SE REPLIENT-ELLES SOUS LEUR CONFORMATION CORRECTE ?	11

PARTIE 1

L'ADN, UNE RÉSERVE D'INFORMATION	15
---	-----------

CHAPITRE 1

LES GÈNES SONT DES UNITÉS MUTABLES	17
LA DÉCOUVERTE DU GÈNE	18
LE RÔLE DES CHROMOSOMES DANS L'HÉRÉDITÉ	22
CHAQUE GÈNE SE SITUE SUR UN CHROMOSOME SPÉCIFIQUE	26
LES GÈNES SONT ALIGNÉS LINÉAIREMENT	27
LA CARTE GÉNÉTIQUE EST CONTINUE	30
UN GÈNE - UNE PROTÉINE	31
UNE DÉFINITION DÉTAILLÉE : LE CISTRON	33
CARTOGRAPHIE DES MUTATIONS AU NIVEAU MOLÉCULAIRE	35

CHAPITRE 2

L'ADN EST LE MATÉRIEL GÉNÉTIQUE	37
DÉCOUVERTE DE L'ADN	37
L'ADN EST LE MATÉRIEL GÉNÉTIQUE (PRESQUE) UNIVERSEL	39
LES COMPOSANTS DE L'ADN	41

L'ADN EST UNE DOUBLE HÉLICE	44
LA RÉPLICATION DE L'ADN EST SEMI-CONSERVATIVE	49
LE CODE GÉNÉTIQUE EST LU SOUS FORME DE TRIPLETS	49
LES MUTATIONS PONCTUELLES CHANGENT UNE SEULE PAIRE DE BASES	52
LES MUTATIONS SONT CONCENTRÉES EN DES POINTS CHAUDS	53
LE TAUX DE MUTATION	55

CHAPITRE 3

LA TOPOLOGIE DES ACIDES NUCLÉIQUES

57

L'ADN PEUT ÊTRE DÉNATURÉ ET RENATURÉ	57
LES ACIDES NUCLÉIQUES S'HYBRIDENT PAR APPARIEMENT DES BASES	58
LES ACIDES NUCLÉIQUES MONO-CATÉNAIRES PEUVENT AVOIR UNE STRUCTURE SECONDAIRE	60
RÉPÉTITIONS INVERSES ET STRUCTURE SECONDAIRE	63
L'ADN DOUBLE-BRIN A DES STRUCTURES EN DOUBLE HÉLICE ALTERNATIVES	65
UNE FORME D'ADN TOURNANT À GAUCHE	66
L'ADN FERMÉ PEUT ÊTRE SUPERENROULÉ	68
LE SUPERENROULEMENT INFLUENCE LA STRUCTURE DE LA DOUBLE HÉLICE	70

CHAPITRE 4

ISOLEMENT DU GÈNE

72

LES ENZYMES DE RESTRICTION CLIVENT L'ADN EN FRAGMENTS SPÉCIFIQUES	73
CONSTRUCTION D'UNE CARTE DE RESTRICTION À PARTIR DES FRAGMENTS	75
LES SITES DE RESTRICTION PEUVENT ÊTRE UTILISÉS COMME MARQUEURS GÉNÉTIQUES	78
OBTENIR LA SÉQUENCE D'ADN	83
LES GÈNES PROCARYOTES ET LES PROTÉINES SONT COLINÉAIRES	86
LES GÈNES DES EUCARYOTES PEUVENT ÊTRE MORCELÉS	87
CERTAINES SÉQUENCES D'ADN CODENT POUR PLUS D'UNE PROTÉINE	89
L'INFORMATION GÉNÉTIQUE PEUT ÊTRE FOURNIE PAR L'ADN OU PAR L'ARN	90
LA PORTÉE DU PARADIGME	93

PARTIE 2

TRANSFORMER LES GÈNES EN PROTÉINES

95

CHAPITRE 5

LA CHAÎNE DE MONTAGE DE LA SYNTHÈSE PROTÉIQUE

97

L'ARN DE TRANSFERT EST UN ADAPTATEUR	98
L'ARN MESSAGER EST TRADUIT PAR LES RIBOSOMES	99
LA SIGNIFICATION DU CODE GÉNÉTIQUE	103
LES SITES D'ACTION DU RIBOSOME	105
L'INITIATION NÉCESSITE LES SOUS-UNITÉS 30S ET DES FACTEURS ACCESSOIRES	106
UN ARN _i INITIATEUR SPÉCIAL INITIE LA CHAÎNE POLYPEPTIDIQUE	109
L'INITIATION CHEZ LES EUCARYOTES NÉCESSITE PLUSIEURS FACTEURS	112
LE FACTEUR D'ÉLONGATION T APORTE UN AMINOACYL-ARN _i DANS LE SITE A	114
LE RIBOSOME AVANCE PAR TRANSLOCATION	116
TERMINAISON : TROIS CODONS ARRÊTENT LA SYNTHÈSE PROTÉIQUE	120

CHAPITRE 6

L'ARN DE TRANSFERT : L'ADAPTATEUR TRADUCTIONNEL

122

LA FEUILLE DE TRÉFLE UNIVERSELLE	123
LA STRUCTURE TERTIAIRE EST EN FORME DE L	123

COMMENT LES SYNTHÉTASES RECONNAISSENT-ELLES LES ARNt ?	126
DISCRIMINATION DANS L'ÉTAPE DE CHARGEMENT	128
LA RECONNAISSANCE CODON-ANTICODON COMPORTE UNE OSCILLATION	129
L'ARNt COMPORTE DE NOMBREUSES BASES MODIFIÉES	132
LES MODIFICATIONS DE BASES PEUVENT CONTRÔLER LA RECONNAISSANCE DES CODONS	135
LES MITOCHONDRIES ONT UN NOMBRE RÉDUIT D'ARNt	135
LES ARNt MUTÉS LISENT DIFFÉRENTS CODONS	137
LES ARNt SUPPESSEURS SONT EN COMPÉTITION VIS-À-VIS DE LEURS CODONS	140
L'ARNt PEUT INFLUER SUR LE CADRE DE LECTURE	142

CHAPITRE 7

LE RIBOSOME : USINE DE LA TRADUCTION

LES RIBOSOMES SONT DES PARTICULES RIBONUCLÉOPROTÉIQUES COMPACTES	144
INTERACTIONS ENTRE LES PROTÉINES RIBOSOMALES ET L'ARNr	145
LA RECONSTRUCTION <i>IN VITRO</i> REPRODUIT L'ASSEMBLAGE <i>IN VIVO</i>	147
L'ASSEMBLAGE DES SOUS-UNITÉS EST LIÉ À LA TOPOLOGIE	150
TOUS LES CONSTITUANTS D'UN RIBOSOME PEUVENT ÊTRE MUTÉS	151
LES RIBOSOMES ONT PLUSIEURS CENTRES ACTIFS	152
LA PRÉCISION DE LA TRADUCTION	154
	156

CHAPITRE 8

L'ARN MESSAGER MATRICE

LES MESSAGES BACTÉRIENS SONT TRANSITOIRES	159
LA PLUPART DES ARNm BACTÉRIENS SONT POLYCISTRONIQUES	161
TRADUCTION DES MESSAGERS POLYCISTRONIQUES	163
UNE DÉFINITION FONCTIONNELLE DE L'ARNm EUCARYOTE	164
LA PUISSANCE DES SYSTÈMES DE TRADUCTION <i>IN VITRO</i>	166
LA PLUPART DES ARNm EUCARYOTES SONT POLYADÉNYLÉS EN 3'	168
TOUS LES ARNm EUCARYOTES ONT UNE COIFFE MÉTHYLÉE À L'EXTRÉMITÉ 5'	169
L'INITIATION PEUT IMPLIQUER L'APPARIEMENT DE BASES ENTRE L'ARNm ET L'ARNr	171
LES PETITES SOUS-UNITÉS PEUVENT MIGRER VERS LES SITES D'INITIATION SUR L'ARNm EUCARYOTE	172
LA SYNTHÈSE PROTÉIQUE EST LIÉE À LA LOCALISATION CELLULAIRE	173

PARTIE 3

CONTRÔLE DE L'EXPRESSION DES GÈNES PAR LA TRANSCRIPTION

181

CHAPITRE 9

LES INTERACTIONS ARN POLYMERASE-PROMOTEUR CONTRÔLENT L'INITIATION

183

LA TRANSCRIPTION EST CATALYSÉE PAR L'ARN POLYMERASE	184
L'ARN POLYMERASE BACTÉRIENNE CONSISTE EN UNE ENZYME « CŒUR » ET LE FACTEUR SIGMA	185
L'ARN POLYMERASE DES EUCARYOTES EST COMPOSÉE DE NOMBREUSES SOUS-UNITÉS	188
LE FACTEUR SIGMA DES BACTÉRIES CONTRÔLE LA LIAISON À L'ADN	189
LES UNITÉS DE TRANSCRIPTION S'ÉTENDENT DES PROMOTEURS AUX TERMINATEURS	192
LES PROMOTEURS COMPORTENT DES SÉQUENCES CONSENSUS	193
L'ARN POLYMERASE PEUT SE LIER AUX PROMOTEURS <i>IN VITRO</i>	195
LA SUBSTITUTION DES FACTEURS SIGMA POURRAIT CONTRÔLER L'INITIATION	200
LES PROMOTEURS DE L'ARN POLYMERASE II SONT SITUÉS EN AMONT DU SITE D'INITIATION	206
LES PROMOTEURS DE L'ARN POLYMERASE II SONT COMPOSÉS DE MULTIPLES PARTIES	209
LES FACTEURS DE TRANSCRIPTION RECONNAISSENT DES SÉQUENCES CONSENSUS PARTICULIÈRES	211

LES « ENHANCERS » (SÉQUENCES ACTIVATRICES) SONT DES ÉLÉMENTS BIDIRECTIONNELS QUI STIMULENT L'INITIATION	213
L'ARN POLYMÉRASE III POSSÈDE UN PROMOTEUR AVAL	216
CHAPITRE 10	
UNE PANOPLIE D'OPÉRONS :	
LE MODÈLE DE L'OPÉRON LACTOSE ET D'AUTRES EXEMPLES	219
L'INDUCTION ET LA RÉPRESSION SONT CONTRÔLÉES PAR DES PETITES MOLÉCULES	220
LES GROUPES DE GÈNES STRUCTURAUX SONT RÉGULÉS D'UNE FAÇON COORDONNÉE	220
LA PROTÉINE RÉPRESSEUR SE FIXE AU NIVEAU DE L'OPÉRATEUR	223
L'OPÉRATEUR AGIT EN <i>CIS</i>	225
COMMENT LE RÉPRESSEUR BLOQUE-T-IL LA TRANSCRIPTION ?	227
LES CONTACTS DANS L'OPÉRATEUR	228
LE RÉPRESSEUR EST UNE PROTÉINE MULTIMÉTRIQUE SE LIANT À L'ADN	229
QUITTER L'ADN ET STOCKER L'EXCÉDENT DE RÉPRESSEUR	231
UN PARADOXE D'INDUCTION	234
LA RÉPRESSION PEUT SE PRODUIRE À DE MULTIPLES LOCI	234
LA DISTINCTION ENTRE UN CONTRÔLE POSITIF ET NÉGATIF	235
LA RÉPRESSION CATABOLIQUE IMPLIQUE UNE RÉGULATION POSITIVE AU NIVEAU DU PROMOTEUR	236
LES PÉRIODES CRITIQUES PROVOQUENT LA RÉPONSE « STRINGENTE »	240
LA TRADUCTION PEUT ÊTRE CONTRÔLÉE D'UNE FAÇON AUTONOME	242
DES PETITES MOLÉCULES D'ARN PEUVENT RÉGULER LA TRADUCTION	244
CHAPITRE 11	
CONTRÔLE AU NIVEAU DE LA TERMINAISON :	
L'ATTÉNUATION ET L'ANTITERMINAISON	248
DEUX MODES DE TERMINAISON DANS <i>E. COLI</i> METTENT EN JEU DES PALINDROMES	249
LA TERMINAISON CHEZ LES EUCARYOTES MET AUSSI EN JEU DES STRUCTURES SECONDAIRES OU DES RÉPÉTITIONS DE U	251
DES STRUCTURES SECONDAIRES ALTERNATIVES CONTRÔLENT L'ATTÉNUATION	252
LA PORTÉE GÉNÉRALE DE L'ATTÉNUATION	256
COMMENT FONCTIONNE LE FACTEUR RHO DANS <i>E. COLI</i> ?	259
L'ANTITERMINAISON DÉPEND DE SITES SPÉCIFIQUES	261
DAVANTAGE DE SOUS-UNITÉS POUR L'ARN POLYMÉRASE ?	267
CHAPITRE 12	
LES CASCADES LYTIQUES ET LA RÉPRESSION LYSOGÉNIQUE	269
LE DÉVELOPPEMENT LYTIQUE EST CONTRÔLÉ PAR UNE CASCADE	270
LES GROUPEMENTS FONCTIONNELS DANS LES PHAGES T7 ET T4	272
LA CASCADE LYTIQUE DE LAMBDA REPOSE SUR L'ANTITERMINAISON	274
LA LYSOGÉNIE EST ASSURÉE PAR UN CIRCUIT AUTONOME	277
LE RÉPRESSEUR EST UN DIMÈRE QUI SE FIXE DE FAÇON COOPÉRATIVE AU NIVEAU DE CHAQUE OPÉRATEUR	279
COMMENT EST ÉTABLIE LA SYNTHÈSE DU RÉPRESSEUR	285
L'ANTIRÉPRESSEUR EST REQUIS POUR L'INFECTION LYTIQUE	287
UN ÉQUILIBRE DÉLICAT : LA LYSOGÉNIE OU LA LYSE	289
PARTIE 4	
LA PERPÉTUATION DE L'ADN	293
CHAPITRE 13	
LE RÉPLICON : UNITÉ DE RÉPLICATION	295
LA RÉPLICATION SÉQUENTIELLE FORME DES YEUX	296
LE GÉNOME BACTÉRIEN EST UN RÉPLICON UNIQUE	298

RÉPLICATION ET CYCLE CELLULAIRE	299
LES CHROMOSOMES EUCARYOTES CONTIENNENT DE NOMBREUX RÉPLICONS	300
ISOLEMENT DES ORIGINES DES RÉPLICONS CHEZ LA LEVURE	303
DIFFÉRENTES STRUCTURES DE RÉPLICATION : ŒIL DE RÉPLICATION, CERCLE ROULANT ET BOUCLE D	303
L'INCOMPATIBILITÉ ENTRE PLASMIDES EST DÉTERMINÉE PAR LE NOMBRE DE COPIES	307

CHAPITRE 14

L'APPAREIL DE RÉPLICATION DE L'ADN

LES ADN POLYMÉRASES, CES ENZYMES QUI FABRIQUENT L'ADN	313
LA SYNTHÈSE DE L'ADN EST SEMI-DISCONTINUE	316
LES FRAGMENTS D'OKAZAKI SONT AMORCÉS PAR DES ARN	317
LA COMPLEXITÉ DE L'APPAREIL DE RÉPLICATION BACTÉRIEN	319
INITIATION DE LA SYNTHÈSE D'UNE MOLÉCULE D'ADN MONOCATÉNAIRE	320
LE MOUVEMENT DU PRIMOSOME	322
INITIATION DE LA RÉPLICATION AUX ORIGINES BICATÉNAIRES	325
L'APPAREIL DE RÉPLICATION DU PHAGE T4	329
LE PROBLÈME DES RÉPLICONS LINÉAIRES	331

CHAPITRE 15

LES MÉCANISMES DE MAINTENANCE DE L'ADN

MODIFICATION ET RESTRICTION	336
LES ACTIVITÉS ALTERNATIVES DES ENZYMES DE TYPE I	338
LA DOUBLE ACTIVITÉ DES ENZYMES DE TYPE III	341
FACE À DES DÉGÂTS DE L'ADN !	342
LES SYSTÈMES D'EXCISION-RESYNTÈSE D'E. COLI	345
LES SYSTÈMES DE RECOMBINAISON-RÉPARATION D'E. COLI	347
UN SYSTÈME SOS CONSTITUÉ DE NOMBREUX GÈNES	348
LES SYSTÈMES DE RÉPARATION DES MAMMIFÈRES	350

PARTIE 5

CONSTITUTION DU GÉNOME DES EUCARYOTES

351

CHAPITRE 16

L'EXTRAORDINAIRE PUISSANCE DE LA TECHNOLOGIE DES RECOMBINANTS D'ADN

353

TOUTE SÉQUENCE D'ADN PEUT ÊTRE CLONÉE DANS UNE BACTÉRIE	354
CONSTRUIRE UN ADN CHIMÉRIQUE	355
COPIER L'ARN ^m EN ADN	358
ISOLER INDIVIDUELLEMENT DES GÈNES À PARTIR DU GÉNOME	359
LA MARCHÉ SUR LE CHROMOSOME	361
LES GÈNES EUCARYOTES PEUVENT ÊTRE EXPRIMÉS DANS DES SYSTÈMES PROCARYOTES	364

CHAPITRE 17

UN CONTINUUM DE SÉQUENCES RENFERME LES GÈNES DE STRUCTURE

367

LE PARADOXE DE LA VALEUR C DÉCRIT DES VARIATIONS DANS LA TAILLE DU GÉNOME	367
LES CINÉTIQUES DE RÉASSOCIATION DÉPENDENT DE LA COMPLEXITÉ DE LA SÉQUENCE	369
LES GÉNOMES EUCARYOTES CONTIENNENT, EN TERME DE SÉQUENCES, DES COMPOSANTS DIFFÉRENTS	371
LA COMPLEXITÉ DE L'ADN NON RÉPÉTITIF PEUT NOUS PERMETTRE D'ESTIMER LA TAILLE DU GÉNOME	372
LES GÉNOMES EUCARYOTES CONTIENNENT DES SÉQUENCES RÉPÉTITIVES	372
L'ADN MODÉRÉMENT RÉPÉTITIF CONSISTE EN DE NOMBREUSES SÉQUENCES DIFFÉRENTES	373

LES MEMBRES DE FAMILLES DE SÉQUENCES RÉPÉTITIVES SE RESSEMBLENT MAIS NE SONT PAS TOUT À FAIT IDENTIQUES	375
LA PLUPART DES GÈNES DE STRUCTURE FONT PARTIE DE L'ADN NON RÉPÉTITIF	377
COMBIEN Y A-T-IL DE GÈNES NON RÉPÉTITIFS EXPRIMÉS ?	378
LES CINÉTIQUES DES RÉACTIONS DIRIGÉES PAR L'ARN PERMETTENT D'ESTIMER LE NOMBRE DE GÈNES	379
LES GÈNES SONT EXPRIMÉS À DES TAUX TRÈS VARIABLES	381
QUELLES SONT LES PARTIES COMMUNES ENTRE LES DIFFÉRENTES POPULATIONS D'ARNm ?	382

CHAPITRE 18

L'ORGANISATION DES GÈNES MORCELÉS

LES GÈNES EXISTENT SOUS TOUTES SORTES DE FORMES ET DE TAILLE	385
LES INTRONS DES GÈNES CODANT POUR LES ARN _r ET LES ARN _t	388
LES JONCTIONS EXON-INTRON ONT UNE SÉQUENCE CONSENSUS	388
L'INTRON D'UN GÈNE PEUT ÉGALEMENT ÊTRE L'EXON D'UN AUTRE GÈNE	389
COMMENT LES GÈNES MORCELÉS ONT-ILS ÉVOLUÉ ?	392

PARTIE 6

DES GROUPES DE SÉQUENCES APPARENTÉES

397

CHAPITRE 19

LES GÈNES DE STRUCTURE APPARTIENNENT À DES FAMILLES DE TAILLES DIFFÉRENTES

399

LES GÈNES DE LA GLOBINE SONT ORGANISÉS EN DEUX GROUPES	400
UN CROSSING-OVER INÉGAL RÉARRANGE LES GROUPES DE GÈNES	402
DE NOMBREUSES THALASSÉMIES SONT DUES À DES CROSSING-OVER INÉGAUX	402
LES GROUPES DE GÈNES SUBISSENT DE CONTINUELLES RÉORGANISATIONS	405
LA DIVERGENCE DE SÉQUENCE DÉFINIT DEUX CATÉGORIES DE SITES DANS L'ADN	406
L'HORLOGE DE L'ÉVOLUTION RETRACE LE DÉVELOPPEMENT DES GÈNES DE LA GLOBINE	408
LES PSEUDOGÈNES SONT LES « VOIES SANS ISSUE » DE L'ÉVOLUTION	409
DE NOMBREUSES PROTÉINES ONT EN COMMUN DES FAMILLES DE GÈNES	411
CERTAINS GROUPES DE GÈNES EN TANDEM CODENT POUR LES HISTONES	412
L'ARN DE TRANSFERT ET L'ARN RIBOSOMAL POSSÈDENT DES GÈNES RÉPÉTÉS	415
LES DEUX GÈNES DE L'ARN _r SONT CONTENUS DANS UNE UNITÉ DE RÉPÉTITION EN TANDEM	416
GÈNES 5S ET PSEUDOGÈNES SONT INTRIGUÉS	418
UN DILEMME DANS L'ÉVOLUTION : COMMENT LES COPIES MULTIPLES ACTIVES SE MAINTIENNENT-ELLES ?	419

CHAPITRE 20

DES GÉNOMES CONFINÉS DANS LES ORGANITES

422

LES GÈNES DES ORGANITES ONT UNE TRANSMISSION NON MENDÉLIENNE	422
LES GÉNOMES DES ORGANITES SONT DES MOLÉCULES CIRCULAIRES D'ADN	423
LES ORGANITES EXPRIMENT LEURS PROPRES GÈNES	424
LE GRAND GÉNOME MITOCHONDRIAL DE LA LEVURE	427
LE GÉNOME MITOCHONDRIAL COMPACT DES MAMMIFÈRES	428
UNE RECOMBINAISON A LIEU DANS (CERTAINS) ADN D'ORGANITE	430
LES RÉARRANGEMENTS DE L'ADN MITOCHONDRIAL DE LEVURE	430

CHAPITRE 21

ORGANISATION DE L'ADN DE SÉQUENCE SIMPLE

432

L'ADN HAUTEMENT RÉPÉTITIF FORME DES SATELLITES	433
LES ADN SATELLITES SONT SOUVENT LOCALISÉS DANS L'HÉTÉROCHROMATINE	434

LES SATELLITES DES ARTHROPODES POSSÈDENT DE TRÈS COURTES UNITÉS DE RÉPÉTITION IDENTIQUES	435
LES SATELLITES DES MAMMIFÈRES SONT COMPOSÉS D'UNITÉS DE RÉPÉTITIONS HIÉRARCHISÉES	435
EN RETRAÇANT LES STADES D'ÉVOLUTION DE L'ADN SATELLITE DE SOURIS	438
LES VARIATIONS DANS L'UNITÉ DE RÉPÉTITION ACTUELLE	439
LES CONSÉQUENCES DU CROSSING-OVER INÉGAL	440
LA FIXATION PAR RECOMBINAISON POURRAIT PERPÉTUER DES RÉPÉTITIONS IDENTIQUES	442

PARTIE 7

VERS LE PRODUIT FINAL : LA MATURATION DES ARN 443

CHAPITRE 22

EXCISION ET RABOTAGE D'UN ARN STABLE 445

L'ARNase III LIBÈRE LES ARNm PRÉCOCES DU PHAGE T7	448
LES CLIVAGES SONT NÉCESSAIRES POUR LIBÉRER LES ARNr PROCARYOTES ET EUCARYOTES	450
LES GÈNES D'ARNt SONT CLIVÉS À PARTIR DE GROUPES DE GÈNES ET RABOTÉS PAR PLUSIEURS ENZYMES	454

CHAPITRE 23

L'ARN EST UN CATALYSEUR : LES MÉCANISMES DE L'ÉPISSAGE 458

L'EXCISION-ÉPISSAGE DE L'ARNt DE LA LEVURE IMPLIQUE COUPURE ET JONCTION	459
L'ARN PEUT POSSÉDER UNE ACTIVITÉ CATALYTIQUE	461
CERTAINS INTRONS MITOCHONDRIAUX PRÉSENTENT DES SIMILITUDES AVEC LES INTRONS DOUÉS D'AUTO-ÉPISSAGE	465
UN INTRON POURRAIT CODER POUR UNE PROTÉINE RÉGULATRICE	468
L'ÉPISSAGE DE L'ARN NUCLÉAIRE SUIT DES PROCESSUS PRÉFÉRENTIELS	472
LES JONCTIONS NUCLÉAIRES D'ÉPISSAGE PEUVENT ÊTRE INTERCHANGEABLES	473
UN « SPLICÉOSOME » NUCLÉAIRE ENGENDRE UN LASSO	475
LES snARN SONT-ILS IMPLIQUÉS DANS LES RÉACTIONS D'EXCISION-ÉPISSAGE ?	478
LES RÉACTIONS D'ÉPISSAGE ONT-ELLES DES POINTS COMMUNS ?	479

CHAPITRE 24

LE CONTRÔLE DE LA MATURATION DE L'ARN 483

L'h ⁿ ARN EST GRAND ET INSTABLE	483
L'ARNm EST DÉRIVÉ DE L'h ⁿ ARN	485
LA POLYADÉNYLATION ET LA FORMATION DES EXTRÉMITÉS 3'	487
Y A-T-IL UN CONTRÔLE APRÈS LA TRANSCRIPTION ?	489
MODÈLES DU CONTRÔLE DE L'EXPRESSION DES GÈNES	491
LES PROPRIÉTÉS DES POLYPROTÉINES CELLULAIRES	494

PARTIE 8

LA CONDENSATION DE L'ADN 497

CHAPITRE 25

AU SUJET DES GÉNOMES ET DES CHROMOSOMES 499

LA CONDENSATION DES GÉNOMES VIRAUX DANS LEUR CAPSIDE	500
LE GÉNOME BACTÉRIEN EST UN NUCLÉOÏDE AYANT DE NOMBREUSES BOUCLES SUPERENROULÉES	503
LE CONTRASTE ENTRE LA CHROMATINE INTERPHASIQUE ET LES CHROMOSOMES MITOTIQUES	507
LES CHROMOSOMES EUCARYOTES COMME MOYEN DE SÉGRÉGATION	509
CERTAINS GÈNES SONT EXTRACHROMOSOMIQUES	512

L'ÉTAT DÉROULÉ DES CHROMOSOMES EN ÉCOUVILLON	513
LA POLYTÉNIE CRÉE DES CHROMOSOMES GÉANTS	515
LA TRANSCRIPTION BOULEVERSE LA STRUCTURE DES CHROMOSOMES	517

CHAPITRE 26

STRUCTURE DE LA CHROMATINE : LE NUCLÉOSOME 519

LES CONSTITUANTS PROTÉIQUES DE LA CHROMATINE	520
LE NUCLÉOSOME EST LA SOUS-UNITÉ DE BASE DE TOUTE LA CHROMATINE	521
LA PARTICULE CŒUR EST HAUTEMENT CONSERVÉE	524
L'ADN EST ENROULÉ AUTOUR DE L'OCTAMÈRE D'HISTONE	525
SUPERENROULEMENT ET PÉRIODICITÉ DE L'ADN	530
LES NUCLÉOSOMES SONT-ILS ORGANISÉS EN PHASE ?	531
L'ORGANISATION DES NUCLÉOSOMES DANS LA FIBRE DE CHROMATINE	533
BOUCLES, DOMAINES ET SQUELETTES	535

CHAPITRE 27

LA NATURE DE LA CHROMATINE ACTIVE 538

ASSEMBLAGE DES NUCLÉOSOMES CONTRE REPRODUCTION DE LA CHROMATINE	539
L'ASSEMBLAGE DES NUCLÉOSOMES NÉCESSITE DES PROTÉINES NON HISTONES	540
LES GÈNES TRANSCRITS SONT-ILS ORGANISÉS EN NUCLÉOSOMES ?	542
LES DOMAINES SENSIBLES À L'ADNase DE LA CHROMATINE APTE À ÊTRE TRANSCRITE	545
LES HISTONES SONT MODIFIÉES TRANSITOIREMENT	546
L'EXPRESSION GÉNÉRIQUE EST ASSOCIÉE À LA DÉMÉTHYLATION	549
LES SITES HYPERSENSIBLES À L'ADNase MODIFIENT LA STRUCTURE DE LA CHROMATINE	551
FLEXIBILITÉ STRUCTURALE DANS L'ADN	555
HYPOTHÈSES À PROPOS DE LA NATURE DE L'ACTIVATION DES GÈNES	557

PARTIE 9

LE GÉNOME DYNAMIQUE : L'ADN EN MOUVEMENT 561

CHAPITRE 28

RECOMBINAISON ET AUTRES MANIPULATIONS TOPOLOGIQUES DE L'ADN 563

LA RECOMBINAISON NÉCESSITE UN ACCOLEMENT D'ADN DOUBLE BRIN HOMOLOGUE	564
CASSURE ET SOUDURE IMPLIQUENT UN DOUBLE BRIN HÉTÉROLOGUE D'ADN	565
LES CASSURES DES DOUBLES BRINS PROVOQUENT-ELLES DES RECOMBINAISONS ?	568
ISOLEMENT D'INTERMÉDIAIRES DE RECOMBINAISON	569
L'ÉCHANGE DE BRIN EST FACILITÉ PAR LA PROTÉINE RecA	571
LA PROTÉINE RecA ET LES CONDITIONS DE RECOMBINAISON	574
LA CONVERSION GÉNÉRIQUE EST RESPONSABLE DE LA RECOMBINAISON INTERALLÉLIQUE	576
MANIPULATION TOPOLOGIQUE DE L'ADN	578
LA GYRASE INTRODUIT DES SUPERTOURS NÉGATIFS DANS L'ADN	580
UNE RECOMBINAISON DONNÉE RECONNAÎT DES SITES SPÉCIFIQUES	582
COUPURES DÉCALÉES ET RESOUDURE DANS LA SÉQUENCE CŒUR	583
L'INVERSION PEUT CONTRÔLER L'EXPRESSION D'UN GÈNE	586

CHAPITRE 29

LES ÉLÉMENTS TRANSPOSABLES CHEZ LES BACTÉRIES 589

LES SÉQUENCES D'INSERTION SONT DES TRANSPOSONS SIMPLES	590
LES TRANSPOSONS COMPOSÉS ONT DES MODULES IS	592

UN SEUL MODULE DE Tn10 EST FONCTIONNEL	594
LES MODULES DE Tn5 SONT PRESQUE IDENTIQUES MAIS TRÈS DIFFÉRENTS	596
RECOMBINAISON CONSERVATIVE CONTRE RECOMBINAISON RÉPLICATIVE	597
LES INTERMÉDIAIRES DE LA TRANSPOSITION	599

CHAPITRE 30

LES ÉLÉMENTS MOBILES CHEZ LES EUCARYOTES 605

LES ÉLÉMENTS DE CONTRÔLE DU MAÏS SONT TRANSPOSABLES	605
Ds PEUT TRANSPOSER OU ENTRAÎNER UNE CASSURE CHROMOSOMIQUE	607
LA TRANSPOSITION DE Ds EST LIÉE À LA RÉPLICATION	609
LE CYCLE DE VIE DES RÉTROVIRUS IMPLIQUE DES ÉVÉNEMENTS ANALOGUES À LA TRANSPOSITION	610
LES RÉTROVIRUS PEUVENT TRANSDUIRE DES SÉQUENCES CELLULAIRES	614
DES TRANSPOSITIONS ARN-DÉPENDANTES ONT PU AVOIR LIEU DANS LA CELLULE	616
LA FAMILLE Alu	617
LES ÉLÉMENTS TY DE LA LEVURE RESSEMBLENT AUX RÉTROVIRUS	618
IL Y A DE NOMBREUX ÉLÉMENTS TRANSPOSABLES CHEZ <i>D. MELANOGASTER</i>	620
LE RÔLE DES ÉLÉMENTS TRANSPOSABLES DANS LA DYSGÉNÉSIE HYBRIDE	622

CHAPITRE 31

MODIFICATIONS DE L'ORGANISATION DU GÉNOME 625

DES VARIATIONS SPÉCIFIQUES D'UN TISSU SE PRODUISENT DANS LE GÉNOME DE <i>DROSOPHILA</i>	626
SÉLECTION DE SÉQUENCES GÉNOMIQUES AMPLIFIÉES	628
DES SÉQUENCES EXOGÈNES PEUVENT ÊTRE INTRODUITES PAR TRANSFECTION	632
L'ADN TRANSFECTÉ PEUT PÉNÉTRER DANS LES CELLULES GERMINALES	634

PARTIE 10

LES GÈNES AU COURS DU DÉVELOPPEMENT 639

CHAPITRE 32

RÉARRANGEMENTS ET CONSTITUTION DE LA DIVERSITÉ IMMUNITAIRE 641

L'ORGANISATION DES IMMUNOGLOBINES	642
LES GÈNES DES IMMUNOGLOBINES SONT ASSEMBLÉS À PARTIR DE LEURS DIFFÉRENTS ÉLÉMENTS	645
LA DIVERSITÉ DE L'INFORMATION GERMINALE	647
LES RÉACTIONS DE JONCTION ENGENDRENT UNE DIVERSITÉ SUPPLÉMENTAIRE	649
LA RECOMBINAISON DES GÈNES V ET C ENGENDRE DES DÉLÉTIONS ET DES RÉARRANGEMENTS	650
QUELQUES CAUSES POSSIBLES D'EXCLUSION ALLÉLIQUE	653
LA RECOMBINAISON CONTINUE DE L'ADN ENTRAÎNE LA COMMUTATION DE CLASSE	654
L'EXPRESSION PRÉCOCE DES CHÂÎNES LOURDES PEUT ÊTRE MODIFIÉE PAR MATURATION DE L'ARN	656
DES MUTATIONS SOMATIQUES ENGENDRENT UNE DIVERSITÉ SUPPLÉMENTAIRE	658
LE RÉCEPTEUR DES CELLULES T EST APPARENTÉ AUX IMMUNOGLOBULINES	659
COMPLEXITÉ DES PRINCIPAUX LOCI D'HISTOCOMPATIBILITÉ	660

CHAPITRE 33

CHANGER L'ORGANISATION GÉNIQUE DU DEDANS ET DU DEHORS 664

LA LEVURE POSSÈDE DES LOCI MUETS ET DES LOCI ACTIFS POUR LE TYPE SEXUEL	664
LES CASSETTES Muettes ET Actives ONT LES MÊMES SÉQUENCES	666
LA TRANSPOSITION UNIDIRECTIONNELLE EST INITIÉE PAR LE LOCUS RÉCEPTEUR MAT	669

LES TRYPANOSOMES RÉARRANGENT L'ADN POUR EXPRIMER L'ANTIGÈNE DE SURFACE	670
L'INTERACTION DE L'ADN DU PLASMIDE TI AVEC LE GÉNOME DE LA PLANTE	675

CHAPITRE 34

LA RÉGULATION DES GÈNES : LE CHANGEMENT DES PROGRAMMES D'EXPRESSION

LOCALISER LES MUTATIONS DANS LES EXONS ET DANS LES INTRONS	682
QUELQUES GÉNÉRALITÉS SUR LE DÉVELOPPEMENT DES <i>DROSOPHILES</i>	685
LES LOCI COMPLEXES SONT EXTRÊMEMENT GRANDS ET INTERVIENNENT DANS LA RÉGULATION	689
UN MOTIF CODANT COMMUN : LA BOÎTE HOME0 (HOME0 BOX)	695

CHAPITRE 35

ONCOGÈNES : EXPRESSION ANORMALE DES GÈNES ET CANCER

LES VIRUS TRANSFORMANTS PEUVENT ÊTRE PORTEURS D'ONCOGÈNES	699
LES ONCOGÈNES RÉTROVIRAUX ONT DES ÉQUIVALENTS CELLULAIRES	701
LES PROTO-ONCOGÈNES RAS PEUVENT ÊTRE ACTIVÉS PAR MUTATION	703
MYC AINSI QUE D'AUTRES ONCOGÈNES PEUVENT ÊTRE ACTIVÉS PAR INSERTION, TRANSLOCATION ET AMPLIFICATION	706
IMMORTALISATION ET TRANSFORMATION	711
LES FONCTIONS POSSIBLES DES ONCOPROTÉINES	712

ÉPILOGUE

DÉCOUVERTES MAJEURES QUI ONT CHANGÉ NOTRE VISION DE LA GÉNÉTIQUE

GLOSSAIRE	721
-----------	-----

INDEX	739
-------	-----

GENES

Des groupes incroyablement petits d'atomes, beaucoup plus petits que ceux
à leur échelle moléculaire voisine, jouent un rôle clé dans la compréhension des actions
et réactions à l'échelle d'un organisme vivant. Les gènes sont les unités de base
pour comprendre le comportement complexe et régulier des organismes.

Ernst Haeckel, 1916.

INTRODUCTION

Des groupes incroyablement petits d'atomes, beaucoup trop petits pour obéir à des lois statistiques précises, jouent en fait un rôle dominant dans les événements très ordonnés et réglementés à l'intérieur d'un organisme vivant... Le gène est beaucoup trop petit... pour occasionner un comportement ordonné et réglementé conformément à la physique.

Erwin Schrödinger, 1945

BCU IAS / CENTRAL UNIVERSITY LIBRARY

LES CELLULES OBÉISSENT AUX LOIS DE LA PHYSIQUE ET DE LA CHIMIE

A partir du moment où nous avons réalisé qu'un organisme ne transmet pas un simulacre de lui-même à la génération suivante mais qu'au contraire il lui fournit un **matériel génétique** contenant l'**information** nécessaire à l'élaboration d'un descendant, nous avons voulu définir la nature de cette information et la manière dont elle est utilisée. Maintenant que nous connaissons la structure physique du matériel génétique, nous pouvons affirmer que l'objet de la biologie moderne est de définir la complexité des organismes vivants en fonction des propriétés de leurs constituants.

Le **gène** est l'unité d'information génétique, et l'idée que c'était une entité distincte fut l'étape importante dans la définition de sa nature. L'ère de la biologie moléculaire du gène a débuté en 1945 quand Schrödinger émit l'idée que les lois de la physique devaient être inadéquates pour rendre compte des propriétés du matériel génétique, en particulier sa stabilité sur un très grand nombre de générations. On s'attendait à ce que le gène obéisse aux lois de la physique déjà énoncées, mais on pensait que la caractérisation du matériel génétique pouvait conduire à la découverte de nouvelles lois de physique, perspective qui amena de nombreux physiciens à la biologie.

Schrödinger a résumé les propriétés prévisibles du gène en termes qui, après une légère redéfinition, correspondent remarquablement bien à notre idée actuelle : « Nous supposons que la structure du gène est celle d'une énorme molécule, uniquement capable d'un chan-

gement discret qui consiste en un réarrangement des atomes et qui conduit à une molécule isomérique. Le réarrangement (mutation) peut toucher seulement une petite région du gène, et un grand nombre de réarrangements différents peut être possible ».

Évidemment, nous savons maintenant que le gène est effectivement une énorme molécule, mais une molécule qui ne fonctionne pas de façon autonome ; au contraire elle dépend d'autres constituants cellulaires pour se perpétuer et fonctionner. Toutes ses activités obéissent aux lois connues de la physique et de la chimie ; et il n'a pas été nécessaire, finalement, d'invoquer de nouvelles règles.

L'HÉRÉDITÉ FONCTIONNE PAR L'ENTREMISE DE MACROMOLÉCULES

Tous les organismes vivants sont composés de cellules, et peuvent être classés en fonction du type des cellules. Les bactéries sont les **procaryotes**, organismes chez lesquels il existe un seul compartiment cellulaire, entouré par une membrane ou des membranes qui le protègent du monde extérieur. Les **eucaryotes** sont définis par la présence dans chaque cellule d'un **noyau** qui contient le matériel génétique, entouré par un **cytoplasme**, qui à son tour est limité par la **membrane plasmique** qui marque la périphérie de la cellule.

Les **organismes unicellulaires** (procaryotes ou eucaryotes) sont des cellules individuelles, capables de survivre, de se reproduire, et (dans certains cas) de s'accoupler. Les **organismes pluricellulaires** (habituellement eucaryotes) existent grâce à la coopération entre de nombreuses cellules, qui peuvent avoir des spécialités différentes pour contribuer à la survie de l'individu.

La nature de l'information génétique et les moyens généraux (bien que différents dans certains détails) de son expression sont les mêmes chez les procaryotes et les eucaryotes. En décrivant comment la transmission de l'information génétique rend compte de la perpétuation des procaryotes et des eucaryotes, nous pouvons considérer leurs instructions génétiques comme étant du même type, bien que le processus de lecture des instructions implique des étapes supplémentaires chez les eucaryotes.

On peut classer les molécules qui constituent les cellules en deux classes :

- Les **petites molécules** qui sont les substrats et les produits des voies métaboliques, et qui fournissent l'énergie nécessaire à la survie cellulaire ; elles se répartissent en quatre classes : **sucres, acides gras, acides aminés et nucléotides**.
- Les molécules plus grandes — les constituants structuraux de la cellule — qui sont synthétisées à partir des petites molécules. Les **polysaccharides** sont assemblés à partir des sucres ; les **lipides** sont assemblés à partir des acides gras ; les **protéines** sont assemblées à partir des acides aminés ; et les **acides nucléiques** sont assemblés à partir des nucléotides.

Les protéines et les acides nucléiques sont de très grandes molécules, ou **macromolécules**. Elles sont responsables de la transmission de l'information génétique. Les acides nucléiques portent l'information, alors que les protéines fournissent le moyen de la transmettre. Dans chaque cas, *la séquence selon laquelle les différents éléments constitutifs sont assemblés est le facteur clé déterminant la propriété de la macromolécule qui en résulte.*

Les caractéristiques physiques et la localisation d'une macromolécule sont importantes. Une cellule, après tout, n'est pas uniquement constituée de protoplasme contenu dans un sac membraneux : chaque élément structural occupe une place spécifique. Ceci ne concerne pas seulement le fait que les membranes contiennent les lipides alors que le noyau contient le matériel génétique, mais aussi le fait que la perpétuation d'une cellule d'une génération à l'autre dépend du maintien d'une structure hautement ordonnée, dans laquelle des fonctions particu-

lières peuvent être exercées uniquement à des places particulières.

Les réactions biologiques sont effectuées par des enzymes, protéines qui possèdent des activités catalytiques spécifiques qui les rendent capables de réunir deux parties dans un environnement favorable au développement d'une réaction chimique. Les enzymes effectuent les réactions selon lesquelles les voies métaboliques convertissent les matières premières en composés dont l'organisme a besoin. Elles sont responsables de la rupture des composés provenant de l'environnement et de la reconstruction des composants nécessaires au maintien de l'organisme.

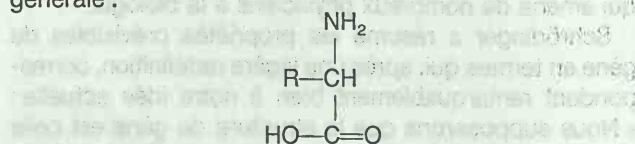
L'assemblage des acides aminés dans une protéine est dirigé par un acide nucléique dont la séquence des nucléotides détermine la séquence des acides aminés. La séquence des acides aminés détermine à son tour la fonction de la protéine. *Toutes les structures macromoléculaires de la cellule sont donc déterminées par la séquence du matériel génétique.*

Est-ce que la structure, en terme macromoléculaire, est suffisante pour expliquer toutes les propriétés de la cellule ? En d'autres termes, la seule structure d'une macromolécule peut-elle expliquer sa localisation dans la cellule, ou une forme externe d'information est-elle nécessaire ? Certaines macromolécules peuvent être capables de trouver leur propre localisation à n'importe quel moment ; d'autres ne peuvent être correctement localisées que pendant leur synthèse.

Au cours du développement d'un organisme multicellulaire à partir d'un œuf fertilisé il est nécessaire que les macromolécules soient localisées à la bonne place. On ne comprend pas très bien cet effet, mais la relation entre localisation et fonction est quelquefois appelée **information de position**, une expression qui implique l'importance de la localisation aussi bien que de la structure elle-même.

LES PROTÉINES SONT DES ENCHAÎNEMENTS D'ACIDES AMINÉS

Chaque protéine est constituée d'une séquence unique d'acides aminés. Un acide aminé libre a la structure générale :



Les acides aminés sont reliés les uns aux autres, pour former une protéine, par des **liaisons peptidiques** résultant

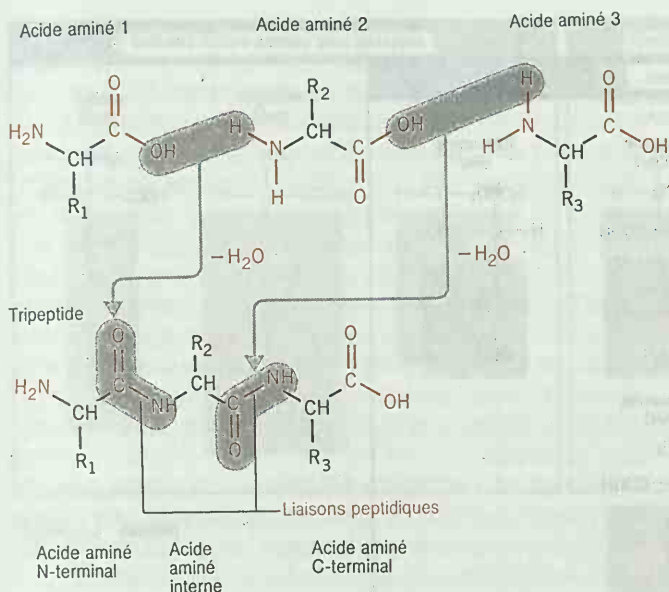


Figure I.1

Une liaison peptidique est formée par une réaction de condensation (appelée ainsi à cause de la perte d'eau)

tant de la *condensation* d'un groupe aminé (NH_2) d'un acide aminé avec le groupe carboxyle ($COOH$) du suivant (Figure I.1). Puisque les liaisons peptidiques sont covalentes, elles sont relativement stables et rarement détruites dans la cellule vivante. Cette destruction résulte habituellement d'une action enzymatique spécifique.

Un **peptide** est constitué d'un petit nombre d'acides aminés réunis par des liaisons peptidiques. Un enchaînement plus long d'acides aminés reliés de cette façon est un **polypeptide**. Le terme de **protéine** est habituellement utilisé pour désigner l'unité fonctionnelle ; elle peut être constituée d'une seule ou de plusieurs chaînes polypeptidiques.

On peut définir la direction d'une chaîne polypeptidique d'après l'orientation des liaisons peptidiques. L'acide aminé situé à une des extrémités de la chaîne doit avoir un groupement NH_2 libre et définit ainsi l'extrémité **amino-** ou **N-terminale**, alors que l'acide aminé de l'autre extrémité doit avoir un groupement $COOH$ libre et définit ainsi l'extrémité **carboxyl-** ou **C-terminale**. Par convention, les séquences de protéine sont écrites de l'extrémité N-terminale (à gauche) vers l'extrémité C-terminale (à droite).

Dans les représentations bidimensionnelles, la chaîne polypeptidique forme un **squelette** en zig-zag, duquel dépassent les **groupements latéraux** (notés **R**). Le groupement **R** est différent pour chaque acide aminé et détermine la nature de sa contribution à la structure générale de la protéine. Il n'y a que vingt acides aminés,

dont la structure est montrée dans la Figure I.2, utilisés pour la synthèse protéique.

Classés selon leurs charges ioniques, les acides aminés forment trois classes. Lysine, arginine et histidine sont basiques. L'addition d'un ion hydrogène convertit le groupement aminé (second) libre de la lysine ou de l'arginine sous la forme NH_3^+ chargée positivement. Un proton peut, de la même manière être ajouté au cycle histidine. L'acide aspartique et l'acide glutamique sont acides, parce que le groupement carboxyle peut perdre un ion hydrogène pour donner la forme $O=C-O^-$ chargée négativement. Les autres acides aminés n'ont aucune charge nette et sont donc neutres.

Certains acides aminés neutres sont polaires (chargés électriquement à cause de la distribution des charges à l'intérieur de la molécule), alors que d'autres sont hydrophobes (excluant l'eau). Les acides aminés polaires (tels que les acides aminés basiques ou acides) peuvent réagir avec d'autres groupements (non protéiques) porteurs de charges électriques. D'autre part, des regroupements d'acides aminés hydrophobes dans une protéine peuvent créer des « poches » qui accueillent des composés non chargés.

En plus des vingt acides aminés « standards », certains autres sont trouvés occasionnellement dans les protéines. Ils sont formés par **modification** d'un des acides aminés standards, après *incorporation de celui-ci dans la protéine*. Ces modifications changent les propriétés du groupement **R** et ainsi jouent souvent des rôles importants

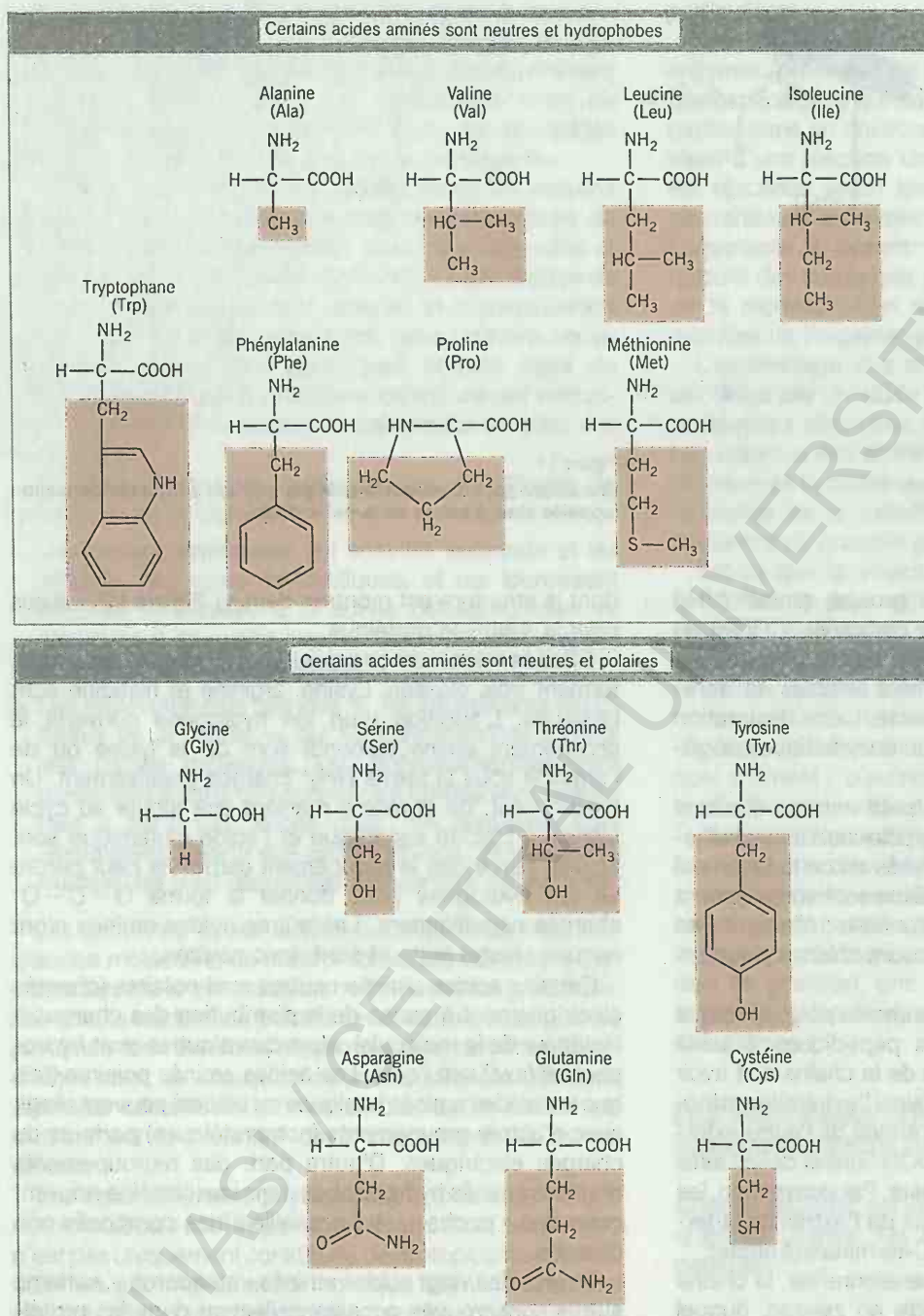


Figure I.2
 Les acides aminés sont classés selon la nature de leur chaîne latérale.
 Le nom de chaque acide aminé est abrégé en trois lettres.

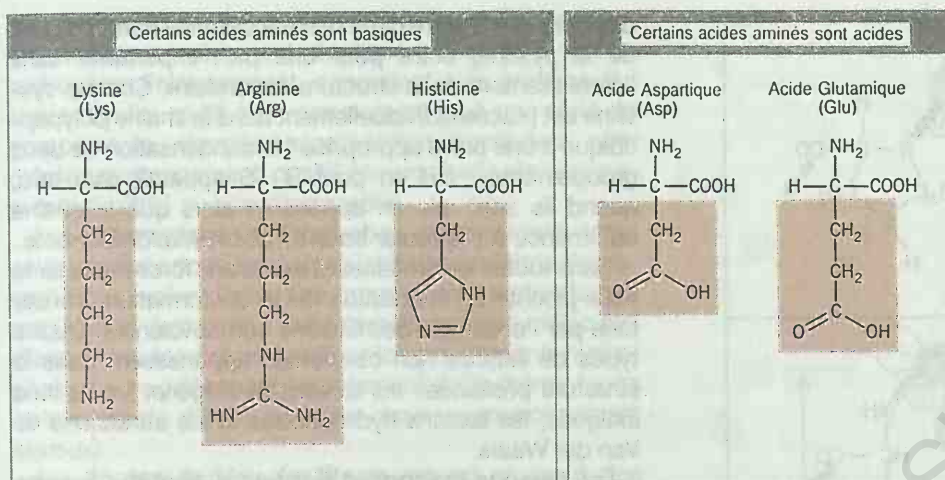


Figure I.2 (suite)

dans la fonction de la protéine. Quelques exemples sont montrés dans la **Figure I.3**.

L'addition d'un petit groupement qui change la charge ionique d'un acide aminé est une des modifications courantes. La **phosphorylation** résulte de l'addition d'un groupement phosphate, habituellement au groupement R de la sérine ou de la tyrosine. Une protéine avec des groupements phosphate est parfois appelée une **phosphoprotéine**. Des charges négatives sont introduites avec le groupement phosphate.

L'**acétylation** et la **méthylation** résultent de l'addition de groupements acétyles ou méthyles habituellement à la lysine, acide aminé basique. Cette addition peut empêcher la formation de charges positives sur le groupement aminé.

Certains acides aminés modifiés sont trouvés dans des protéines ou des types de cellules spécifiques. Par exemple, l'hydroxyproline et l'hydroxylysine sont des dérivés de la proline et de la lysine qui ont reçu un groupement $-\text{OH}$ supplémentaire ; on les trouve dans les collagènes, protéines du tissu conjonctif.

Un changement de structure plus important apparaît dans les protéines **glycosylées**, où une chaîne latérale glucidique est reliée à un acide aminé. De telles protéines sont appelées **glycoprotéines** ; beaucoup de protéines qui fonctionnent dans un compartiment extracellulaire comportent des chaînes latérales oligosaccharidiques très longues. On distingue les **N-oligosaccharides** les plus courants, où le sucre est lié au groupement aminé de l'asparagine, et les **O-oligosaccharides** moins cou-

rants, où le sucre est lié au groupement hydroxyle de la sérine ou de la thréonine.

Ce ne sont en aucun cas les seules modifications qui apparaissent dans les protéines, bien qu'elles soient les plus importantes. Toute modification est importante car son existence étend encore plus le répertoire des structures et des fonctions que les protéines peuvent montrer. Il faut également noter qu'il n'est pas obligatoire qu'une partie associée soit liée à la protéine de façon covalente ; beaucoup de protéines sont associées à des groupements supplémentaires qui participent à leur structure ou à leur fonction, mais qui ne sont pas reliés de façon covalente à un acide aminé.

LES FORCES NON COVALENTES SONT IMPORTANTES DANS LA CONFORMATION DES PROTÉINES

Une caractéristique importante d'une protéine est sa capacité à se replier en une **conformation** tridimensionnelle. Chaque protéine existe sous une conformation unique (ou parfois une série de conformations alternées). La conformation est définie par différents niveaux structuraux.

L'ensemble des acides aminés liés dans la chaîne polypeptidique constitue sa **structure primaire**. La **structure secondaire** est engendrée par l'enroulement de la séquence primaire ; cet enroulement est rendu possible par une libre rotation autour de certaines liaisons. La structure secondaire définit la façon dont le squelette polypeptidique de la protéine se dispose dans l'espace.

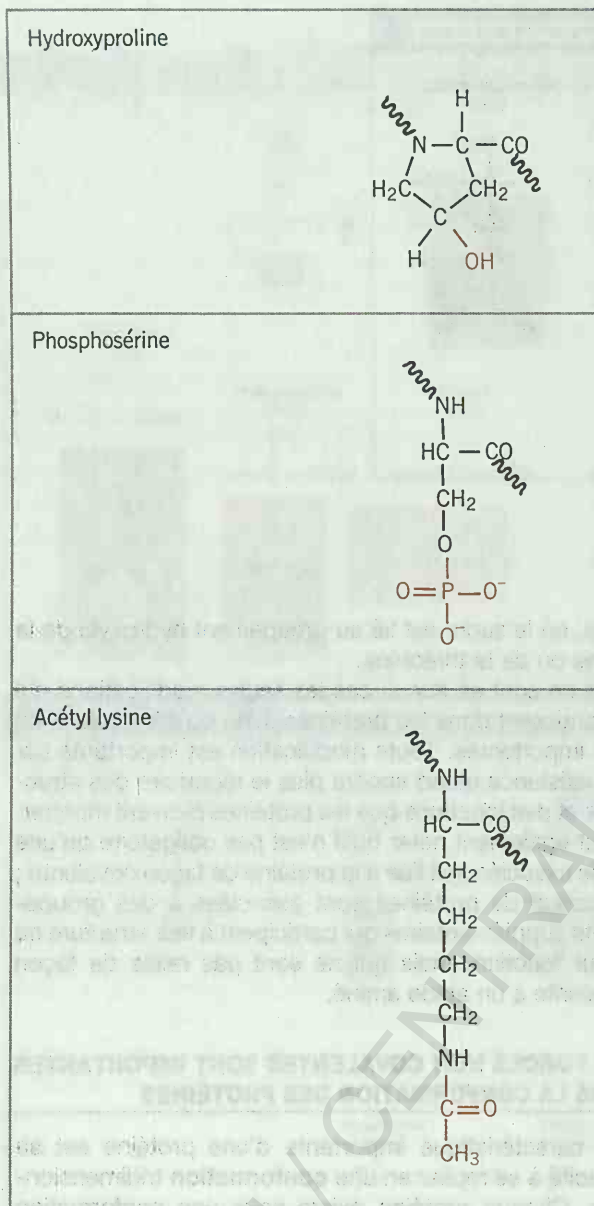


Figure 1.3

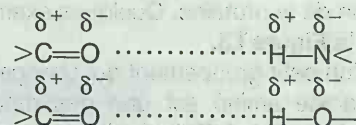
Les acides aminés modifiés sont produits par addition de groupements supplémentaires aux parties réactives des chaînes latérales R des acides aminés qui ont déjà été incorporés dans les protéines.

Plusieurs types d'interactions entre les acides aminés contribuent à la formation de la structure secondaire. Des liaisons covalentes et non covalentes sont impliquées. Dans de nombreuses protéines, la formation de « ponts »

disulfure S – S entre deux résidus cystéine (formant ainsi de la cystine) entre pour une part importante dans l'établissement de la structure secondaire. Chaque cystéine est placée individuellement dans la chaîne polypeptidique à une place appropriée ; la condensation de deux groupements – SH en pont S – S apparaît plus tard, quand ils sont mis en apposition alors que la chaîne commence à s'enrouler sous sa conformation correcte.

Dans toutes les protéines, il existe une force importante sous-jacente à l'acquisition de la conformation, constituée par l'ensemble des liaisons non covalentes. Quatre types de liaisons non covalentes apparaissent dans la structure protéique : les liaisons hydrogène, les liaisons ioniques, les liaisons hydrophobes et les attractions de Van der Waals.

Les **liaisons hydrogène** (liaisons H) sont des liaisons électrostatiques faibles qui se forment entre un atome d'oxygène partiellement chargé négativement et un atome d'hydrogène partiellement chargé positivement, comme dans les exemples



Le δ indique le caractère partiel de la charge électrique. La polarisation des liaisons C = O, N – H ou O – H permet en effet la formation d'une liaison hydrogène entre les deux groupes. La liaison tire son nom du fait que l'atome d'hydrogène est, à un certain degré, partagé entre les groupements fonctionnels.

Les liaisons hydrogène sont beaucoup plus faibles que les liaisons covalentes, d'un facteur supérieur à 10. Comparée à l'énergie de – 50 – 100 kcal/mole libérée par une liaison covalente, l'énergie d'une liaison hydrogène typique est environ – 3 – 6 kcal/mole. (L'énergie d'une liaison est mesurée en kilocalories libérées par mole ; elle est exprimée en valeur négative, comme décrit au Chapitre 3.)

Puisque de nombreuses liaisons hydrogène peuvent être formées dans une macromolécule, leur contribution globale à la stabilité de la conformation peut être importante. Mais la faiblesse d'une liaison hydrogène individuelle (et des autres liaisons non covalentes) rend leur rupture relativement facile dans des conditions physiologiques. Ceci constitue un aspect important de la fonction des protéines et des acides nucléiques.

Les **interactions ioniques** s'établissent entre des groupements de charge opposée (tels que les acides aminés acides et basiques). Dans un environnement aqueux, les groupements chargés réagissent avec l'eau,

qui masque la charge. Ceci réduit la force d'interaction à un niveau sensiblement identique à celle de la liaison hydrogène. Fréquemment, cependant, une protéine qui (par exemple) se lie à un groupement ou à une série de groupements chargés négativement possèdera elle-même un groupement positif (ou une série de groupements positifs) au site d'interaction.

Les **interactions hydrophobes** apparaissent entre acides aminés dont les chaînes latérales apolaires s'agrègent pour exclure l'eau. Les groupements hydrophobes ont un effet destructeur des interactions entre molécules d'eau ; l'eau tend donc à obliger les groupements hydrophobes à se juxtaposer (exactement de la même façon qu'une goutte d'huile se forme dans un environnement aqueux).

Les **forces de Van der Waals** entrent en jeu à des distances très réduites à chaque fois que deux atomes sont attirés l'un vers l'autre. Les forces individuelles sont très faibles, mais elles peuvent devenir significatives dans des régions fortement tassées d'une macromolécule, ou quand deux surfaces macromoléculaires sont en contact étroit. Les forces de Van der Waals individuelles peuvent avoir une force de $-0,5 - 1,0$ kcal/mole.

LES STRUCTURES PROTÉIQUES SONT EXTRÊMEMENT DIVERSES

Les protéines ont une très grande diversité de forme, ceci résultant de leur capacité à produire une vaste gamme de conformations. Certains types de structure secondaire sont relativement communs. Souvent, plusieurs régions organisées en un type particulier de structure secondaire peuvent s'entasser les unes près des autres.

Les liaisons hydrogène établies entre les groupements d'une même chaîne polypeptidique peuvent amener le squelette à se tordre en forme d'hélice, le plus souvent sous la forme **hélice α** illustrée dans la **Figure 1.4**. L'hélice est en partie stabilisée par les liaisons hydrogène formées entre le groupement $C=O$ d'une liaison peptidique et le groupement $-NH$ de la liaison peptidique située quatre résidus plus loin sur la chaîne polypeptidique. L'hélice α est un élément commun des structures secondaires des protéines.

Une chaîne polypeptidique peut être étendue en une structure de type feuillet grâce à des liaisons hydrogène établies avec une autre chaîne dirigée en sens inverse. Une fois encore, la liaison hydrogène implique le groupement $C=O$ d'une liaison peptidique et le groupement NH d'une autre. La structure en **feuillet β** est illustrée dans la **Figure 1.5** ; elle est favorisée par la présence de

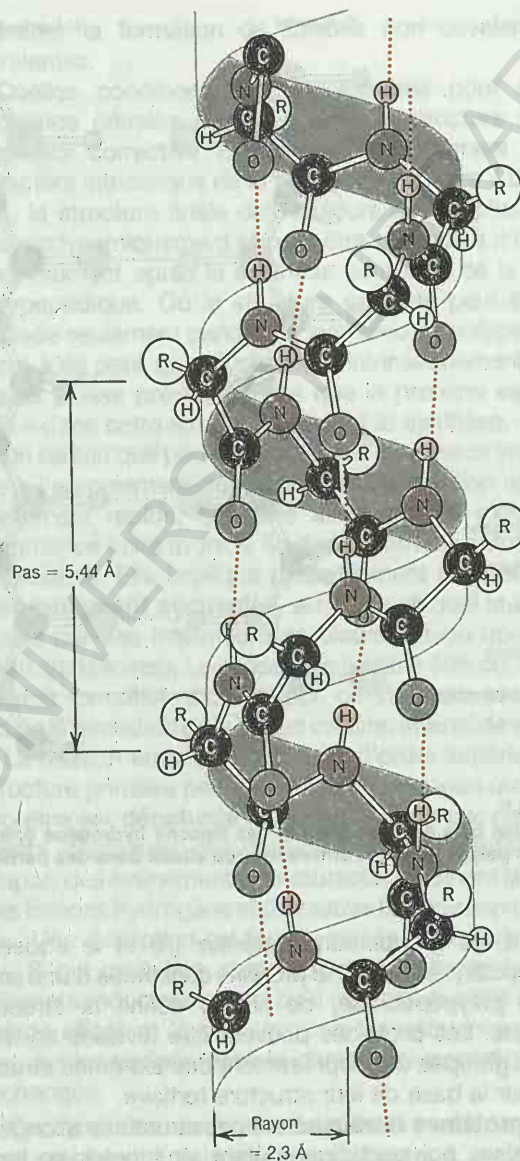


Figure 1.4

Une hélice α polypeptidique est stabilisée par des liaisons hydrogène entre acides aminés s'établissant tous les quatre résidus dans le même polypeptide (Les liaisons H sont indiquées en pointillés).

résidus glycine et alanine. Ce type de structure peut être formé aussi bien entre deux chaînes polypeptidiques différentes qu'entre deux parties d'une même chaîne polypeptidique disposées de telle sorte que les régions adjacentes soient orientées en sens inverse.

La **structure tertiaire** décrit l'organisation tridimensionnelle de tous les atomes de la chaîne polypeptidique,

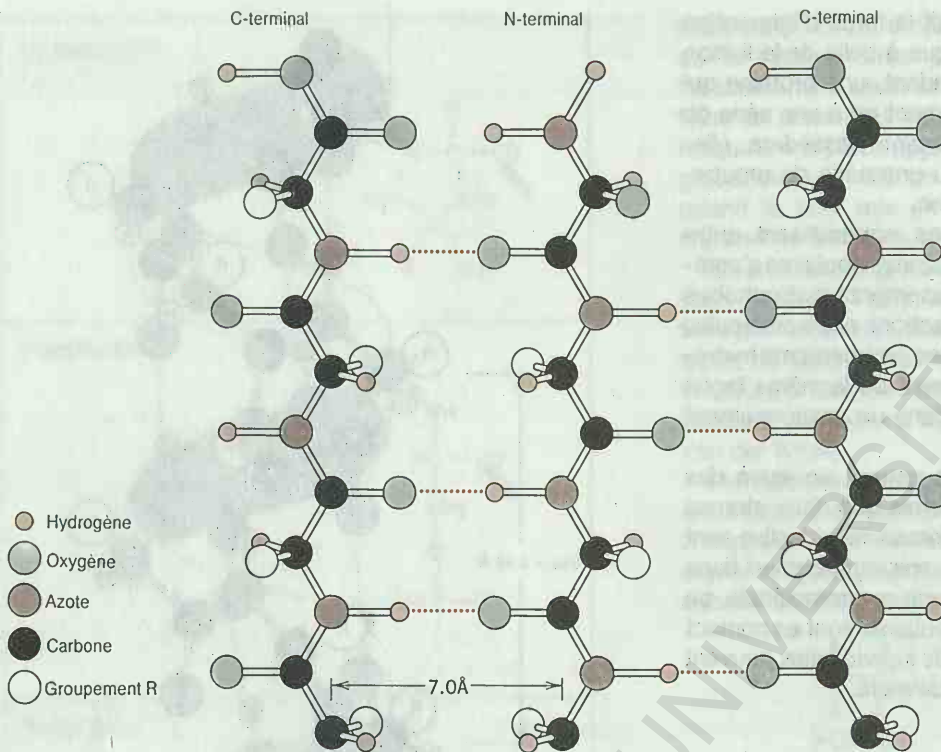


Figure 1.5

Un feuillet bêta est stabilisé par des liaisons hydrogène entre des acides aminés situés dans des chaînes polypeptidiques différentes (ou situés dans des parties de la même chaîne dirigées en sens opposé).

incluant les groupements latéraux (R) et le squelette polypeptidique. Dans une protéine constituée d'une seule chaîne polypeptidique, ce niveau définit la structure complète. Les protéines peuvent être divisées en deux grands groupes (qui représentent des extrêmes structuraux) sur la base de leur structure tertiaire.

Les **protéines fibreuses** ont des structures allongées, les chaînes polypeptidiques étant structurées en longs cordons. Selon les protéines, la structure tertiaire peut reposer sur une hélice α ou sur un feuillet β . La forme particulière de la structure secondaire peut s'étendre à une grande partie de la protéine. Les protéines fibreuses sont les principaux constituants structuraux des cellules et des tissus ; par exemple, elles sont dominantes dans les tissus conjonctifs. Ainsi, elles jouent un rôle statique en fournissant un cadre structural.

Les **protéines globulaires** ont des structures plus compactes. La structure tertiaire est en général plutôt irrégulière, et contient souvent de nombreux types de structure secondaire. Par exemple, bien qu'une partie de

la structure soit souvent sous la forme d'hélice α , il est rare que celle-ci englobe la structure entière. De nombreuses protéines globulaires sont constituées de structures secondaires partiellement hélicoïdales et partiellement non hélicoïdales. Des portions relativement courtes d'hélice α peuvent être maintenues dans une relation particulière l'une avec l'autre. Les protéines globulaires incluent la plupart des enzymes et la plupart des protéines qui vont nous intéresser et qui sont impliquées dans l'expression et la régulation du gène. Leurs rôles tendent à être dynamiques ; ils concernent leur capacité à catalyser des réactions ou à changer de conformation.

Il y a encore un niveau plus élevé d'organisation reconnu dans les **protéines polymériques**, constituées d'agréats de plusieurs chaînes polypeptidiques. La **structure quaternaire** représente la structure globale prise par une protéine oligomérique. Les chaînes polypeptidiques individuelles qui composent une protéine oligomérique sont souvent définies comme étant les **sous-unités de la protéine**.

On peut illustrer la variabilité de la structure protéique en considérant les diverses formes possibles d'une chaîne polypeptidique de 300 acides aminés. Considérée comme une rangée d'acides aminés totalement étendue, elle s'étirerait sur 100 nm ; enroulée en hélice α elle s'étendrait sur 45 nm. Mais organisée en feuillet β , elle deviendrait une boîte plate de 7×7 nm de seulement 0,8 nm de profondeur ; et sous une forme sphérique qui utiliserait diverses formes de structure secondaire, elle aurait un diamètre de 4,3 nm seulement.

Des conformations successives telles que celles-ci sont interconvertibles seulement par rotation autour de liaisons covalentes qui change la position des atomes individuels. Le passage d'une conformation à une autre requiert alors la mise en place et la destruction de liaisons non covalentes seulement, toutes les liaisons covalentes restant intactes. En principe, chaque protéine peut acquérir un nombre presque illimité de conformations, mais en pratique, une ou quelques conformations sont favorisées. Dire qu'une protéine existe dans une conformation donnée, ne signifie pas que la position de chaque atome est figée : il existe un certain degré de mobilité autour d'une position « moyenne » qui caractérise la conformation.

La relation entre structure et fonction dépend du type de protéine. Certaines protéines, spécialement les enzymes qui catalysent les réactions métaboliques, peuvent avoir un **site actif** où la réaction catalytique se réalise. La structure de ce site, résultant de la juxtaposition d'un groupe d'acides aminés, peut être absolument cruciale et n'admet aucune variation. Il peut être possible de modifier d'autres régions de la protéine sans supprimer sa fonction catalytique. Cependant d'autres protéines, par exemple certaines protéines structurales, peuvent avoir une conformation impliquée en grandes parties ou même toute entière directement dans leur fonction qui peut donc être altérée par le moindre changement de conformation. Les protéines diffèrent donc dans le degré avec lequel leur structure peut varier sans perturber leur fonction.

COMMENT LES PROTÉINES SE REPLIENT-ELLES SOUS LEUR CONFORMATION CORRECTE ?

On peut considérer comme principe fondamental que les structures d'ordre supérieur sont déterminées directement par les structures d'ordre inférieur. Ceci signifie que la séquence primaire des acides aminés renferme l'information nécessaire à l'enroulement de la protéine sous la bonne conformation. La réaction d'enroulement peut

entraîner la formation de liaisons non covalentes et covalentes.

Quelles conditions sont nécessaires pour que la séquence primaire s'enroule en une structure d'ordre supérieur correcte ? Est-ce que l'enroulement est un caractère intrinsèque de la séquence primaire ? Dans ce cas, la structure finale doit toujours être la plus stable thermodynamiquement et peut être obtenue à n'importe quel moment après la synthèse complète de la chaîne polypeptidique. Ou la structure correcte peut-elle être réalisée seulement pendant la synthèse du polypeptide ? Alors il se peut qu'une structure intrinsèquement moins stable puisse prévaloir parce que la protéine est « piégée » dans cette structure pendant la synthèse.

On ne sait que peu de choses des processus impliqués dans l'enroulement des protéines. La réaction est habituellement rapide, quelques secondes ou moins. Elle commence juste avant la fin de la synthèse complète de la protéine. Elle implique probablement un mécanisme d'**enroulement séquentiel**, au cours duquel la réaction passe par des intermédiaires discrets (bien que hautement transitoires). Le processus semble être coopératif ; ainsi la formation d'une région de structure secondaire active la formation de la région voisine, et ainsi de suite.

La relation entre les structures d'ordre supérieur et la structure primaire peut être mise en évidence quand une protéine est **dénaturée** par chauffage ou par des traitements chimiques qui détruisent sa conformation. La plupart des événements dénaturants entraînent la rupture des liaisons hydrogène et des autres liaisons non covalentes. Une exception est fournie par la rupture des ponts S-S qui peut être obtenue par le traitement à l'aide d'agents réducteurs. Quoi qu'il en soit, tous ces changements affectent la conformation ; la séquence primaire des acides aminés dans la chaîne polypeptidique reste inchangée.

Dans certains cas, la structure d'ordre supérieur résulte inéluctablement de la séquence primaire. L'enzyme ribonucléase en est l'exemple classique. Après dénaturation, la conformation active de la protéine peut être restituée en inversant le processus de dénaturation. *Toute l'information nécessaire pour former la structure secondaire réside dans la séquence primaire.* Ainsi la production d'une ribonucléase active apparaît inévitablement chaque fois que la chaîne primaire est placée dans les conditions appropriées.

Dans d'autres cas, les protéines peuvent être irréversiblement dénaturées. Ainsi sous certaines conditions (non physiologiques), une protéine peut avoir d'autres conformations stables. Il est possible (bien que non prouvé) que

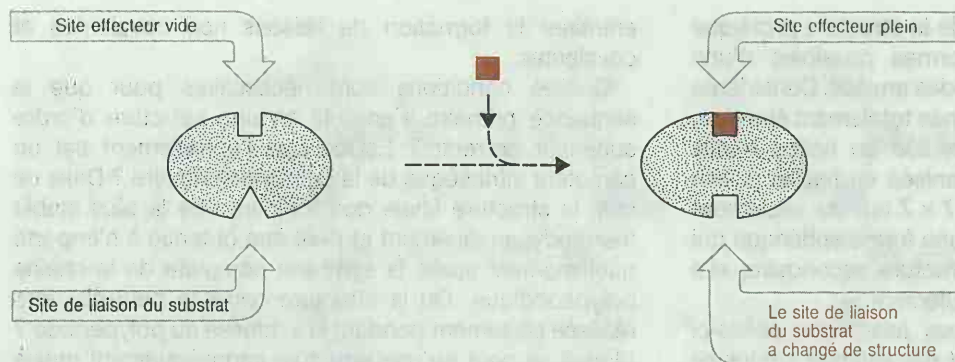


Figure 1.6

Quand une petite molécule se fixe au site effecteur, une protéine allostérique subit une transition vers une autre conformation dans laquelle la structure et la fonction du second site, correspondant au site de fixation du substrat, sont altérées.

la conformation correcte puisse être acquise *uniquement* pendant la synthèse de la protéine ; la conformation pourrait dépendre d'interactions spécifiques entre des régions de la protéine qui peuvent apparaître uniquement en l'absence d'autres régions (c'est-à-dire de celles qui n'ont pas encore été synthétisées).

Les protéines sont synthétisées seulement dans le sens N-terminal vers C-terminal (voir plus bas) et il se pourrait que cette orientation soit importante dans la détermination des structures d'ordre supérieur. De même, comme il a été précédemment souligné, certaines protéines peuvent trouver leur propre localisation cellulaire uniquement pendant leur synthèse.

Dans certains cas, un **cofacteur** qui est une partie de la protéine active (tel que le groupe hème liant le fer des cytochromes) doit être présent pour que la chaîne polypeptidique prenne sa conformation correcte. Dans le cas des protéines oligomériques, il peut être nécessaire qu'une sous-unité soit présente pour qu'une autre acquiert la conformation appropriée.

On en est donc resté à l'idée que la séquence primaire d'une protéine est un déterminant crucial de ses structures d'ordre supérieur ; quelquefois elle peut être en fait le seul déterminant, mais dans d'autres cas des interactions supplémentaires peuvent être impliquées dans l'acquisition de la conformation finale. Dans chaque cas, cependant, si la séquence primaire est synthétisée dans l'environnement approprié, elle recevra inévitablement les superstructures appropriées.

Bien qu'une superstructure résulte de la séquence primaire, la même structure tertiaire générale peut être déterminée par des séquences primaires différentes. Par

exemple, les globines (globule rouge) de différentes espèces varient substantiellement dans leur séquence, mais acquièrent la même structure tertiaire globale.

Il est important de savoir qu'une protéine peut être constituée de **domaines**. Un domaine est une région de la protéine (relativement) indépendante ; dans certains cas, sa conformation peut être acquise indépendamment par le fragment correspondant de la chaîne polypeptidique. Certaines protéines globulaires sont constituées de domaines discontinus connectés par des « crevasses ».

Un domaine peut représenter une unité fonctionnelle ; une activité de la protéine peut être identifiée à ce domaine particulier. Un domaine peut représenter une unité d'évolution ; il peut être apparu comme un polypeptide fonctionnel ou comme une région de polypeptide et s'être associé plus tard avec d'autres domaines pour donner naissance à une nouvelle protéine dont les capacités sont plus larges. Quelquefois un substrat se lie à la crevasse située entre les domaines.

Les protéines **allostériques** existent dans des conformations alternées et ont des propriétés biologiques différentes dans chaque conformation. La transition entre les conformations peut être influencée par l'interaction de la protéine avec un cofacteur ou avec une autre protéine.

La **Figure 1.6** illustre les conséquences d'une transition allostérique. Le changement de conformation induit par une interaction au niveau d'un site de liaison (le site effecteur) peut modifier la structure et donc la fonction d'un autre site de liaison (le site de liaison du substrat ou site actif). Les transitions allostériques affectent *uniquement* la conformation — elles ne modifient pas

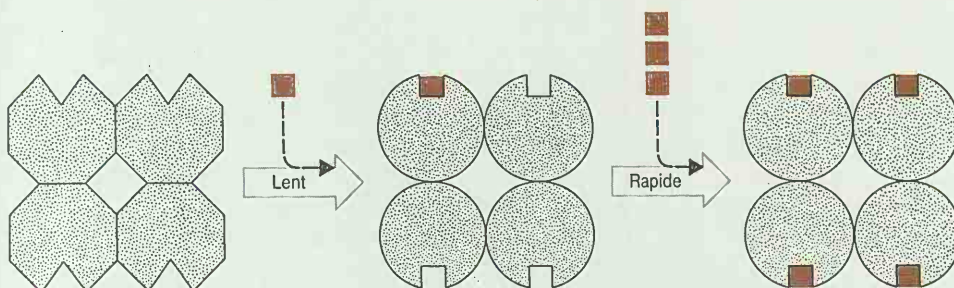


Figure I.7

Quand une protéine allostérique est un oligomère constitué de sous-unités identiques, la fixation sur une sous-unité peut accroître la fixation sur les autres sous-unités.

la séquence primaire — et elles sont étroitement dépendantes de la formation et de la rupture des liaisons hydrogène.

Les protéines allostériques jouent un rôle important dans les régulations métaboliques et génétiques. Souvent, le produit d'une voie métabolique peut se lier à une enzyme qui catalyse une étape initiale de cette voie afin de bloquer le flux des petites molécules à ce niveau. Cette interaction est appelée **rétroinhibition** (feedback inhibition). Dans la régulation génétique, l'activité d'une protéine contrôlant l'expression d'un gène peut répondre à une petite molécule. Dans chaque cas, l'interaction critique est représentée par la capacité d'une petite molécule à se lier à un site spécifique de la protéine, et ainsi à changer la conformation de façon à ce que l'activité de la protéine soit modifiée.

Certaines protéines allostériques sont des oligomères constitués de sous-unités identiques. Dans ce cas, la conformation d'une sous-unité peut modifier la conformation des autres sous-unités. Par exemple, quand une petite molécule qui contrôle l'activité de la protéine se

fixe sur une sous-unité, il peut être plus facile pour les autres sous-unités de fixer la petite molécule. Cet effet est illustré dans la **Figure I.7**. Il amplifie l'effet de liaison de la première petite molécule de telle façon que la protéine bascule de façon caractéristique et très rapidement d'un état à l'autre. Ceci est une propriété importante pour une protéine régulatrice.

Les activités structurales, catalytiques et régulatrices des protéines d'une cellule sont responsables, directement ou indirectement, de la mise en place de son ultrastructure et de l'interconversion des petites molécules. Les protéines elles-mêmes sont synthétisées selon les instructions des acides nucléiques qui constituent l'appareil génétique. Obéissant aux lois de la physique et de la chimie, toutes ces macromolécules sont responsables de la survie des cellules vivantes. Il n'y a pas besoin de postuler des interactions supplémentaires, bien que des questions restent en suspens sur la relation qui existe entre les propriétés des macromolécules et leurs localisations topographiques, c'est-à-dire sur la nature de l'information de position.



Figure 1. Schematic representation of the mechanism of the reaction of the substituted benzene with the substituted benzene. The structures are labeled with 'R' and 'X' to indicate substituents.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

The reaction of the substituted benzene with the substituted benzene is a complex process involving several steps. The first step is the formation of a complex between the two molecules, followed by the transfer of a proton from one molecule to the other, and finally the formation of a product.

CHAPITRE 1

LES GÈNES SONT DES UNITÉS MUTABLES

PARTIE 1

L'ADN, UNE RÉSERVE D'INFORMATION

Quand l'alcool atteint une concentration d'environ 90 %, il y a séparation d'une substance fibreuse qui, au cours de l'agitation du mélange, s'enroule d'elle-même autour de la baguette de verre comme un fil sur une bobine, les autres impuretés restant sous forme de précipité granuleux. Le matériel fibreux est redissout et le processus répété plusieurs fois. En bref, cette substance est hautement réactive et après analyse élémentaire se conforme *très* étroitement aux valeurs théoriques de l'ADN pur (qui aurait pu le deviner).

Oswald Avery, 1944

BCU IASI / CENTRAL UNIVERSITY LIBRARY

PARTIE I

L'ADN, UNE RESERVE D'INFORMATION

Quand l'ADN est considéré comme un simple support d'information, on se contente de le stocker dans des banques de données. Mais si l'on considère l'ADN comme une véritable réserve d'information, on se rend compte qu'il faut lui attribuer une valeur d'information. Cette valeur d'information est fonction de la quantité d'information qu'il contient et de la façon dont cette information est organisée. C'est pourquoi, dans ce chapitre, nous allons nous intéresser à la façon dont l'ADN est organisé et à la façon dont il est utilisé.

Gérard Auvry, 1984

CHAPITRE 1

LES GÈNES SONT DES UNITÉS MUTABLES

Les caractères héréditaires sont définis par leur capacité à être transmis d'une génération à l'autre d'une façon prévisible. Avant de considérer la nature de l'unité d'hérédité, il est important de bien comprendre la distinction entre l'apparence de l'organisme (ce que nous observons) et la constitution génétique sous-jacente (que nous devons déduire). Les propriétés visibles ou autrement mesurables constituent le **phénotype**, alors que les facteurs génétiques responsables de la mise en place du phénotype constituent le **génotype**.

Étant donné leurs effets sur le phénotype, les caractères héréditaires sont très variés dans leur complexité. Certains semblent avoir en principe un effet relativement limité — par exemple, les êtres humains peuvent avoir les yeux marron ou bleus. Certains sont apparemment plus complexes ; par exemple, l'hérédité de la forme d'un nez. A travers le concept de gène, la génétique trouve un terrain commun pour l'hérédité des caractères, depuis la capacité d'effectuer une simple réaction métabolique jusqu'à la construction d'une forme complexe.

Le gène est l'unité d'hérédité. *Chaque gène est une séquence d'acide nucléique qui porte l'information représentant un polypeptide spécifique.* Un gène est une entité stable, mais il peut occasionnellement subir un changement dans sa séquence. Un tel changement est appelé une **mutation**. Quand une mutation se produit, la nouvelle forme du gène est transmise d'une manière stable, exactement comme la forme précédente.

L'organisme portant le gène altéré est un **mutant** ; un organisme portant le gène normal (non altéré) est dit de

type sauvage. « Type sauvage » peut être utilisé pour décrire le génotype ou le phénotype. Puisque la plupart des mutations endommagent la fonction d'une protéine, l'étude des mutations est détournée vers l'examen des situations dans lesquelles le gène ne fonctionne pas correctement.

Les effets phénotypiques d'un changement dans une protéine dépendent des fonctions de celle-ci dans l'organisme ; ces effets peuvent être nuls ou léthaux ou évoluer entre ces deux extrêmes. Des changements complexes dans le phénotype peuvent résulter de mutations uniques. Un exemple classique de mutation néfaste qui exerce son effet en interférant avec une voie métabolique a été découvert par Garrod au début du siècle. Son expression « erreurs innées du métabolisme » porte le concept qu'un défaut génétique peut causer une défaillance métabolique qui produira un phénotype mutant.

La phénylcétonurie est une maladie héréditaire de l'homme qui, parmi d'autres effets, perturbe la fonction cérébrale. La maladie résulte de l'absence de l'enzyme qui convertit habituellement l'acide aminé phénylalanine en un autre acide aminé, la tyrosine. Le défaut conduit à l'accumulation de la phénylalanine, qui est alors transformée en un composé toxique, l'acide phénylpyruvique, et toutes sortes de troubles s'ensuivent.

Comme le révèlent les effets des mutations, certains caractères phénotypiques sont déterminés par des gènes uniques, alors que d'autres sont déterminés par plusieurs gènes. Ainsi certaines caractéristiques sont modifiées seulement quand un gène spécifique est muté, alors que

d'autres peuvent être touchées par une mutation sur n'importe lequel des différents gènes. Quand un caractère phénotypique peut être associé à la fonction d'une protéine spécifique, on peut rattacher le phénotype et le génotype (comme, par exemple, dans le cas de la phénylcétonurie).

De nombreux caractères phénotypiques complexes n'ont pas encore de base connue (par exemple, la

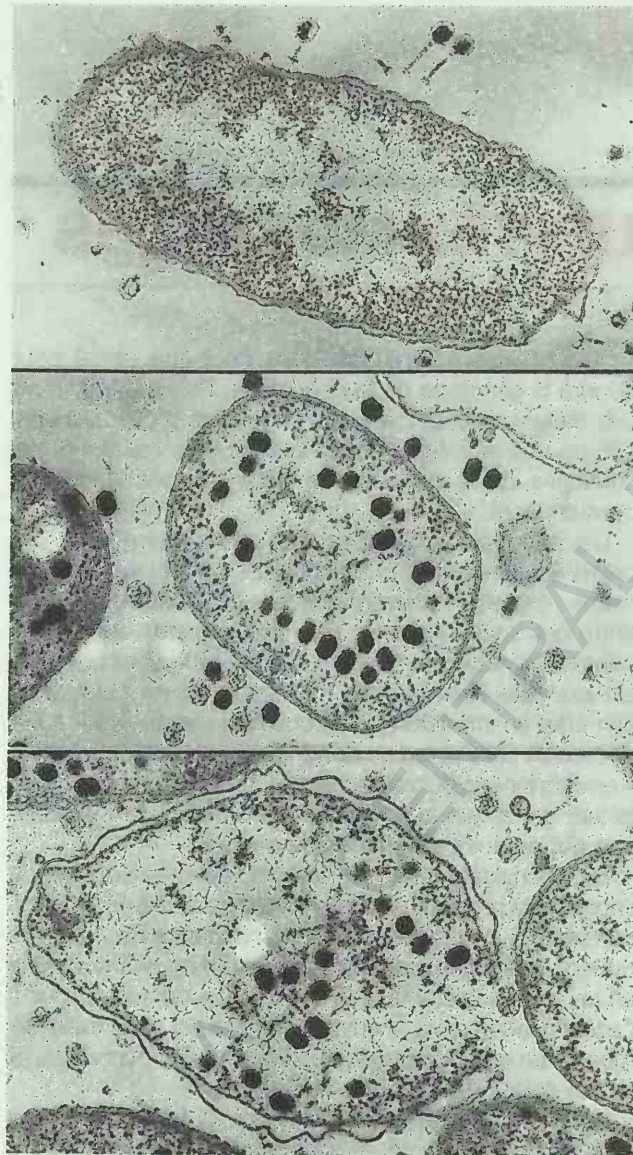


Figure 1.1
L'hérédité des virus se poursuit à travers un cycle au cours duquel un virus infecte une cellule hôte pour donner un virus identique à lui-même.

détermination de la forme d'un nez), et dans ces cas notre seul fil conducteur est constitué par l'hypothèse de travail selon laquelle les interactions des protéines représentées par des gènes non encore identifiés expliqueront les différences phénotypiques entre les individus.

Le concept que le génotype détermine le phénotype s'applique indifféremment aux organismes vivants et aux virus. Les virus prennent la forme physique de particules excessivement petites. Ils ont en commun avec les organismes qu'une génération succède à une autre ; ils en diffèrent par leur incapacité de se propager par eux-mêmes puisqu'ils ont besoin d'infecter une **cellule hôte**.

Les cellules procaryotes et eucaryotes sont sujettes aux infections virales ; les virus qui infectent les bactéries sont habituellement appelés **bactériophages** ou **phages**. Un exemple d'infection phagique est montré dans la **Figure 1.1**. Les virus diffèrent dans leur complexité et dans leurs effets sur une cellule infectée, mais ils possèdent la propriété générale de produire, au cours de l'infection, de nouvelles particules virales qui ressemblent à celles qui ont infecté la cellule hôte au départ.

Le phénotype d'un virus est représenté non seulement par la structure de la particule virale elle-même, mais également par son effet sur la cellule infectée. Des changements dans le génotype du virus peuvent modifier soit la particule elle-même, soit le phénotype de la cellule infectée. De la même façon que les cellules et les organismes, les souches virales peuvent être séparées en type sauvage ou mutant, leur type est hérité par leur descendance, et un type peut être transformé de façon stable en un autre par mutation. L'hérédité des virus a donc les mêmes caractéristiques que l'hérédité des organismes vivants.

L'hérédité virale est portée par des gènes ; la nature et le fonctionnement des gènes viraux sont identiques à ceux des gènes cellulaires. La cellule propage un ensemble assez complexe de gènes d'une manière autonome ; un virus propage un ensemble de gènes plus petit de manière non autonome.

LA DÉCOUVERTE DU GÈNE

Les propriétés essentielles du gène ont été définies par Mendel voici plus d'un siècle. Si l'on résume les deux lois de Mendel, le gène est « un facteur particulaire » que les parents transmettent sans changement à leur descendance. Nous considérons généralement la génétique du point de vue de l'organisme, mais l'alternance des géné-

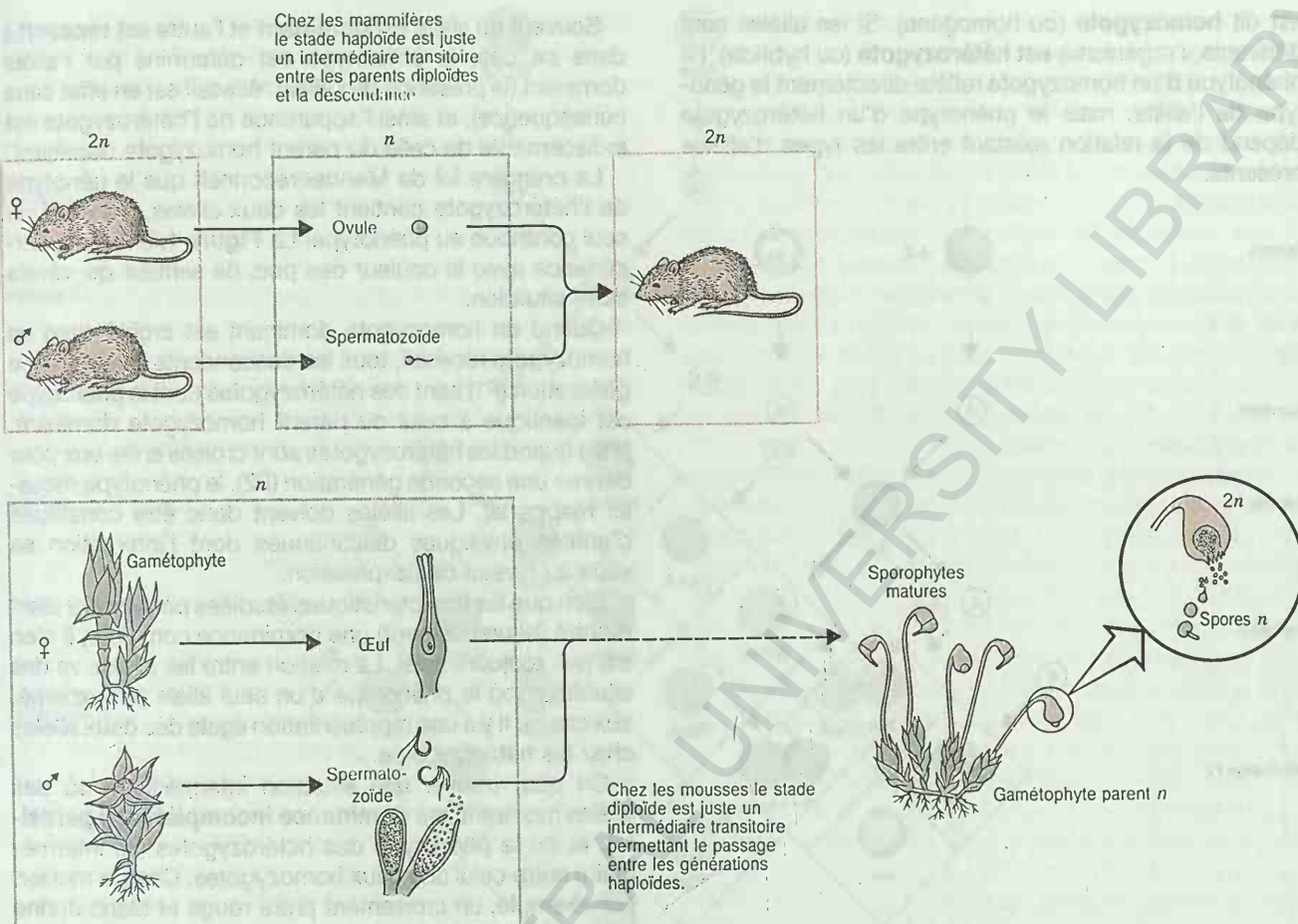


Figure 1.2

Quand des eucaryotes perpétuent leurs gènes à travers une alternance d'états diploïdes et haploïdes, un état fournit le type prédominant alors que l'autre est représenté dans les gamètes. Les mammifères et les mousses sont des exemples extrêmes chez lesquels les générations prédominantes sont respectivement diploïdes et haploïdes.

Le rouge indique le tissu diploïde ($2n$) ; le gris indique le tissu haploïde (n).

illustrée dans la **Figure 1.2** reconnaît le sens dans lequel l'organisme constitue le moyen d'expression et de préservation du gène.

La figure illustre le cycle vital des organismes qui passent par un stade **diploïde**, au cours duquel il y a deux copies de chaque gène. Une des deux copies est transmise par un parent au **gamète** (cellule sexuelle : œuf ou spermatozoïde). Les différents types de gamètes produits par des parents de sexe différent s'unissent pour former un **zygote** (œuf fertilisé). Ainsi, le zygote reçoit une copie de chaque gène de chaque parent, restaurant la situation dans laquelle tout organisme diploïde possède

une copie d'origine paternelle et une copie d'origine maternelle.

Un gène peut exister sous des formes alternatives qui déterminent l'expression d'une caractéristique particulière (par exemple, la couleur d'une fleur peut être rouge ou blanche). Ces formes sont appelées des **allèles**. La première loi de Mendel décrit la **ségrégation des allèles** : les allèles n'ont aucun effet permanent l'un sur l'autre quand ils sont présents dans la même plante, mais ségrègent non modifiés en passant dans les différents gamètes.

Quand les deux allèles sont identiques, un organisme

est dit **homozygote** (ou homogène). Si les allèles sont différents, l'organisme est **hétérozygote** (ou hybride). Le phénotype d'un homozygote reflète directement le génotype de l'allèle, mais le phénotype d'un hétérozygote dépend de la relation existant entre les types d'allèles présents.

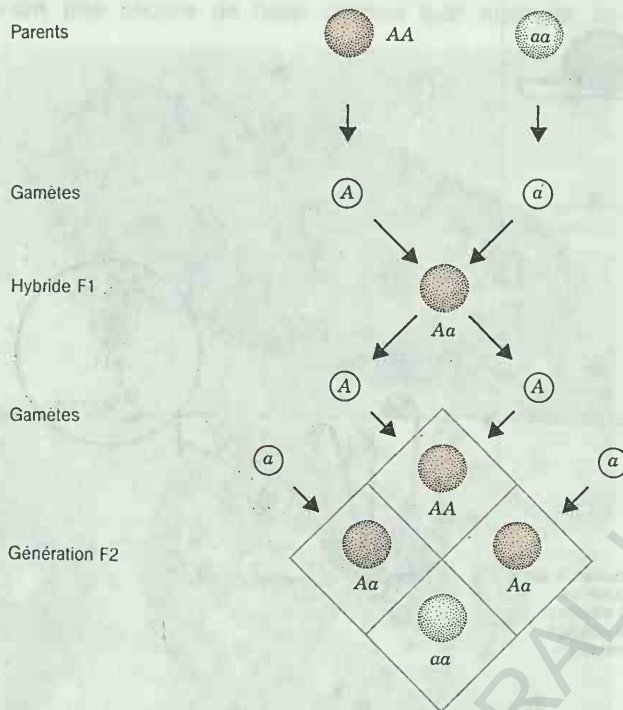


Figure 1.3

Les allèles ségrègent à chaque génération.

Les deux parents sont homozygotes. AA a deux copies de l'allèle dominant ; aa a deux copies de l'allèle récessif. Chaque parent forme seulement un type de gamète, de sorte que la génération F1 est uniformément hybride Aa. Puisque A est dominant sur a, le phénotype de Aa est le même que celui de AA (indiqué en couleur). Le phénotype de l'homozygote récessif aa est indiqué par l'absence de couleur.

Les prédictions de la première loi de Mendel

Chaque hybride F1 forme les gamètes A et a en quantités égales. À l'accouplement, ceux-ci s'unissent au hasard pour une génération F2 constituée de 1AA : 2Aa : 1aa. Puisque AA et Aa ont le même phénotype, on obtient le rapport classique dominant : récessif de 3 : 1.

Dans les cas où l'hétérozygote Aa possède un phénotype intermédiaire entre les AA et aa parentaux, la F1 sera distincte de chaque parent ; et à la F2 le rapport des phénotypes sera 1 dominant : 2 intermédiaires : 1 récessif.

Souvent un allèle est **dominant** et l'autre est **récessif** ; dans ce cas, le phénotype est déterminé par l'allèle dominant (la présence de l'allèle récessif est en effet sans conséquence), et ainsi l'apparence de l'hétérozygote est indiscernable de celle du parent homozygote dominant.

La première loi de Mendel reconnaît que le génotype de l'hétérozygote contient les deux allèles, même si un seul contribue au phénotype. La **Figure 1.3** illustre l'expérience avec la couleur des pois de senteur qui révéla cette situation.

Quand un homozygote dominant est croisé avec un homozygote récessif, tous les descendants à la première génération (F1) sont des hétérozygotes dont le phénotype est identique à celui du parent homozygote dominant. Mais quand les hétérozygotes sont croisés entre eux pour donner une seconde génération (F2), le phénotype récessif réapparaît. Les allèles doivent donc être constitués d'entités physiques discontinues dont l'interaction se situe au niveau de l'expression.

Bien que les caractéristiques étudiées par Mendel aient montré (heureusement) une dominance complète, il n'en est pas toujours ainsi. La relation entre les allèles va des situations où le phénotype d'un seul allèle est exprimé, aux cas où il y a une représentation égale des deux allèles chez les hétérozygotes.

On peut trouver une situation intermédiaire où des allèles montrent une **dominance incomplète (ou partielle)**, et où le phénotype des hétérozygotes est intermédiaire entre celui des deux homozygotes. Chez le muflier, par exemple, un croisement entre rouge et blanc donne naissance à des hétérozygotes à fleurs roses. Bien qu'il n'y ait aucune dominance, on observe la même règle avec une première génération hybride (F1) possédant un phénotype uniforme ; et les mêmes rapports sont conservés à la seconde génération (F2), sauf que l'on peut observer trois phénotypes au lieu de deux.

Les allèles sont dits **codominants** lorsqu'ils contribuent de façon équivalente au phénotype. Par exemple, dans les groupes sanguins humains, les combinaisons AA et BB sont homozygotes, et AB est un hétérozygote codominant chez lequel les groupes A et B sont exprimés de façon équivalente.

La seconde loi de Mendel résume l'**assortiment indépendant des différents gènes**. Quand un homozygote, dominant pour *deux caractères différents*, est croisé avec un homozygote récessif pour les deux caractères, la F1 est constituée comme précédemment de plantes dont le phénotype est identique à celui du parent dominant. Mais à la génération suivante (F2), on observe deux catégories

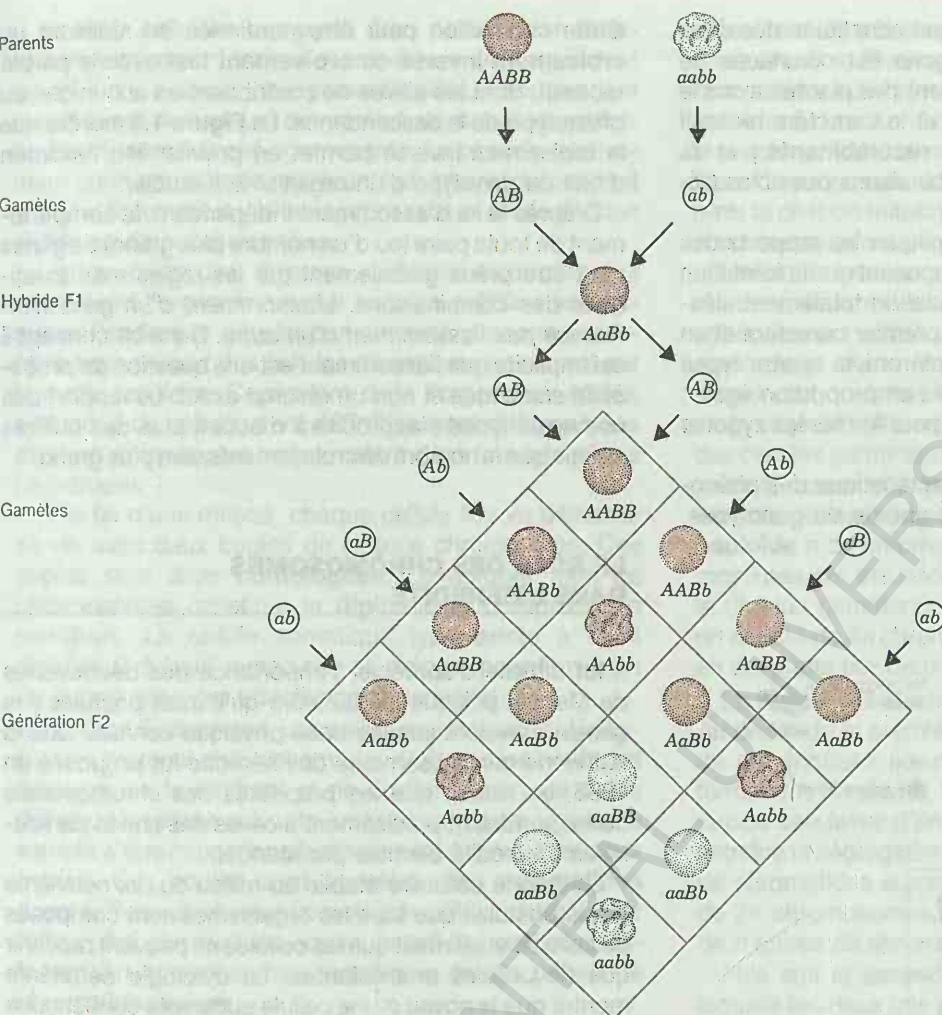


Figure 1.4
Des gènes différents se répartissent indépendamment.

Un parent est homozygote pour deux gènes dominants, A déterminant la couleur et B déterminant la forme (structure ronde). L'autre parent est homozygote pour les allèles récessifs a et b (pas de couleur et forme plissée). La F1 est uniforme pour les caractères dominants.

Les prédictions de la seconde loi de Mendel

Dans chaque plante F1, les allèles ségrègent et les gènes s'assortissent indépendamment, de telle sorte qu'il apparaît un nombre égal de chacun des quatre types possibles de gamète. Les gamètes s'unissent au hasard pour former 9 classes génotypiques. A cause de la dominance de A et de B , il y a seulement quatre classes phénotypiques : 9 colorées lisses : 3 colorées plissées : 3 incolores lisses : 1 incolore plissée.

Notons que chaque génotype réciproque est présent en nombre égal ; par exemple, les deux parents (un $AABB$ et un $aabb$) ou n'importe quel type de recombinant (tel qu'un $AAbb$ et un $aaBB$). Les rapports 3 : 1 sont maintenus pour chaque caractère individuel qui ségrège.

Le nombre de classes phénotypiques sera plus grand si un ou deux des caractères ne sont pas dominants (de sorte que les hétérozygotes apparaissent différents de chaque homozygote). Il sera inférieur si deux gènes affectent un seul caractère du phénotype. Ainsi si A et B sont nécessaires pour produire la couleur, la F2 aura un rapport de 9 colorés : 7 incolores.

de descendants. Une catégorie est constituée des deux **types parentaux**. L'autre catégorie est constituée de **nouveaux** phénotypes, représentant des plantes ayant le caractère dominant d'un parent et le caractère récessif de l'autre. Ce sont des **types recombinants** ; et ils apparaissent avec les deux combinaisons possibles (**ré-ciproques**).

Avec la **Figure 1.4** on peut expliquer les rapports des quatre phénotypes de la F2 en supposant que la formation des gamètes implique une association totalement aléatoire entre un des allèles pour le premier caractère et un des allèles pour le second caractère. Les quatre types de gamètes possibles sont formés en proportion égale ; ensuite ils s'associent au hasard pour former les zygotes de la génération suivante.

Encore une fois, le rapport caractéristique des phénotypes dissimule une variété plus grande de génotypes.

Cette conclusion peut être confirmée en réalisant un **croisement inverse** ou **croisement test** avec le parent récessif, dont les allèles ne contribuent en aucun cas au phénotype de la descendance. La **Figure 1.5** montre que le croisement inverse permet en premier lieu l'examen direct du génotype d'un organisme à étudier.

D'après la loi d'assortiment indépendant, le comportement de toute paire (ou d'un nombre plus grand) de gènes peut être prédit globalement par les règles mathématiques des combinaisons. L'*assortiment d'un gène n'influence pas l'assortiment d'un autre*. Dans ce concept il est implicite que l'assortiment est une question de *probabilité statistique* et non un résultat exact. Le rapport des nouveaux types s'approchera d'autant plus des chiffres prévus que le nombre de croisements sera plus grand.

LE RÔLE DES CHROMOSOMES DANS L'HÉRÉDITÉ

Il fut difficile d'apprécier l'importance des découvertes de Mendel puisque les facteurs qu'il avait postulés (les gènes) n'avaient aucune base physique connue. Quand la théorie chromosomique de l'hérédité fut proposée en 1903, on réalisa que les propriétés des chromosomes correspondaient précisément à celles des unités particulières d'hérédité décrites par Mendel.

La théorie cellulaire établie au milieu du dix-neuvième siècle postulait que tous les organismes sont composés de cellules et affirmait que les cellules ne peuvent provenir que de cellules préexistantes. La cytologie débutante montra que le noyau d'une cellule eucaryote contient une région granuleuse, la **chromatine**, que l'on peut reconnaître grâce à sa réaction avec certains colorants. La chromatine est constituée d'un certain nombre de particules filiformes, les **chromosomes**.

Dans la plupart des cellules, les chromosomes ne peuvent être visualisés que pendant le processus de la division cellulaire. Les deux types de division observés chez les organismes à reproduction sexuée, expliquent la perpétuation du matériel génétique à l'intérieur d'un organisme et le processus d'hérédité prédit par les lois de Mendel.

Un organisme débute sous la forme d'un œuf fertilisé, puis se développe par de nombreuses divisions cellulaires, chaque division donnant deux cellules à partir d'une. Le **cycle cellulaire** englobe la période séparant la libération d'une cellule en tant que cellule fille et sa propre division en deux autres cellules filles.

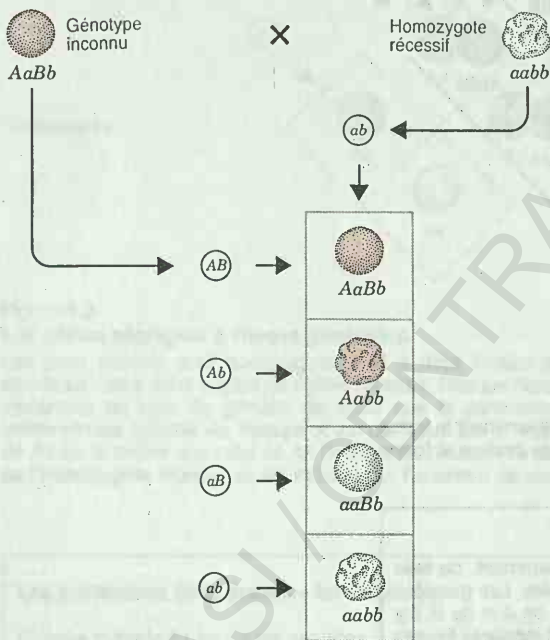


Figure 1.5

Le test de croisement avec un homozygote récessif est utilisé pour définir les génotypes.

Des gamètes sont formés à partir d'un génotype inconnu selon les lois de Mendel. Le génotype de chaque gamète est révélé directement par le résultat de sa combinaison avec un gamète qui apporte seulement les allèles récessifs.

Le cycle cellulaire comporte deux parties. Une **interphase** relativement longue représente le temps pendant lequel il n'y a aucun changement visible, mais pendant lequel la cellule (unique) entreprend ses activités synthétiques et reproduit ses composants. La période relativement courte de **mitose** représente un intermède pendant lequel s'accomplit le véritable processus exact de division visible en deux cellules filles.

Les produits des diverses divisions mitotiques qui constituent l'organisme entier sont les **cellules somatiques**. Pendant le développement embryonnaire, beaucoup ou la plupart des cellules somatiques passeront par le cycle cellulaire. Cependant dans l'organisme adulte, beaucoup de cellules sont **définitivement différenciées** et ne se divisent plus ; elles restent dans une interphase perpétuelle.

À la fin d'une mitose, chaque cellule fille va démarrer sa vie avec deux copies de chaque chromosome. Ces copies sont dites **homologues**. Le nombre total de chromosomes constitue la **diploïdie** et comprend $2n$ membres. La cellule somatique type existe à l'état diploïde ($2n$) (sauf quand elle se prépare pour la mitose ou quand elle y est effectivement engagée).

Pendant l'interphase, une cellule en croissance dédouble son matériel chromosomique. Cette action n'est pas évidente sur le moment et ne devient apparente qu'au début de la mitose suivante, quand chaque chromosome semble s'être coupé longitudinalement pour donner deux copies. Ces copies sont appelées des **chromatides sœurs**. Puisque chaque homologue a été dédoublé pour former des chromatides sœurs, la cellule contient maintenant $4n$ chromosomes, organisés en $2n$ paires de chromatides sœurs.

L'utilisation du terme « chromosome » est ambiguë. Dans la nomenclature cytologique classique, « chromosome » fait référence à l'entité qui possède le **centromère**, une région responsable de son mouvement à la mitose. Quand le matériel chromosomique est doublé pour former les chromatides sœurs, chaque paire de chromatides sœurs a un seul centromère. Pendant la mitose, le centromère est fonctionnellement dédoublé, et à ce moment chaque chromatide sœur est considérée comme un chromosome à part entière. Il ne faudrait pas laisser ce point de vue traditionnel sur le centromère cacher le point principal, c'est-à-dire en terme de génétique la cellule possède quatre copies de chaque chromosome quand elle entre en mitose. Nous considérerons la division de ce point de vue.

La **Figure 1.6** illustre la suite des événements de la division mitotique. Les chromatides sœurs sont d'abord

tirées vers les pôles opposés de la cellule, de sorte que chaque cellule fille reçoit un membre de chaque paire de chromatides sœurs (maintenant devenue un chromosome individuel). Les $4n$ chromosomes présents au début de la division ont été divisés en deux lots de $2n$ chromosomes. Ce processus est répété au cycle cellulaire suivant. Ainsi la division mitotique assure la constance de l'effectif chromosomique dans les cellules somatiques.

L'effectif chromosomique de toute cellule, visualisé au moment de la mitose, est appelé **caryotype**. Le terme est également utilisé pour désigner la constitution chromosomique d'une espèce : par exemple, le caryotype humain est constitué de 46 chromosomes.

Un objectif différent est poursuivi avec la production des cellules germinales (les gamètes : œufs ou spermatozoïdes), accomplie par un autre type de division. La **méiose** donne des cellules qui contiennent un nombre **haploïde n** de chromosomes. Une fois encore, les chromosomes ont été dédoublés avant que le processus de la division commence. La cellule entre en méiose avec un effectif de $4n$ chromosomes, qui sont ensuite séparés en deux lots par deux divisions successives.

La **Figure 1.7** illustre les deux divisions de la méiose. Au début de la première division, les paires homologues de chromatides sœurs s'unissent ou s'apparient pour former des **bivalents**. *Chaque bivalent contient les quatre copies cellulaires d'un chromosome*. La première division entraîne la ségrégation de chaque bivalent en deux paires de chromatides sœurs. Ceci engendre deux ensembles de $2n$ chromosomes, chaque ensemble étant constitué de n paires de chromatides sœurs.

Puis suit la seconde division méiotique au cours de laquelle les deux lots de $2n$ se divisent à nouveau. Cette division est formellement identique à la division mitotique, puisqu'un membre de chaque paire de chromatides sœurs s'isole dans une cellule fille différente.

Le résultat global de la méiose est donc la répartition des $4n$ chromosomes de départ dans quatre cellules haploïdes (n). Ces cellules peuvent ensuite donner des œufs ou des spermatozoïdes matures.

Au cours de la formation des gamètes, les homologues d'origines paternelle et maternelle sont séparés, de sorte que chaque gamète reçoit *un seul des deux homologues* de son parent. Quand on rapproche ce processus des prédictions des lois de Mendel on observe que les *chromosomes non homologues se séparent indépendamment*. Ainsi n'importe quel membre d'une paire d'homologues entre au hasard dans le gamète avec n'importe quel membre d'une paire d'homologues différente.

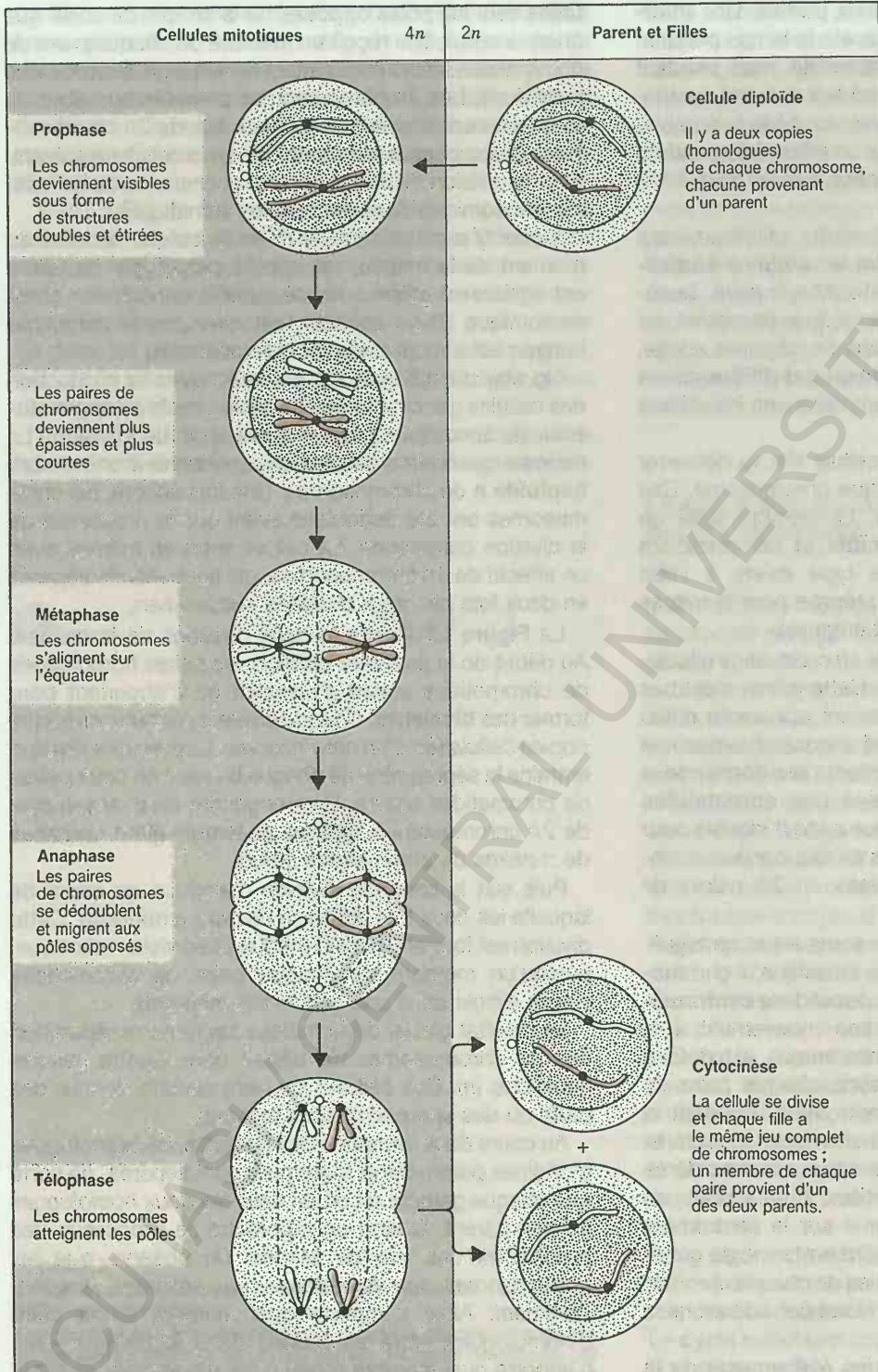


Figure 1.6

La mitose perpétue la constitution chromosomique de la cellule.

La figure montre le comportement d'une paire de chromosomes homologues (en fait une cellule eucaryote en possède de nombreuses paires). L'origine de chaque homologue est indiquée par sa couleur (l'homologue blanc provient d'un parent, l'homologue coloré de l'autre parent). Chaque chromosome est dupliqué avant le début de la mitose. Pendant la mitose, les doubles se séparent et sont isolés dans des cellules différentes. Chaque cellule fille a le même effectif chromosomique que la cellule parentale.

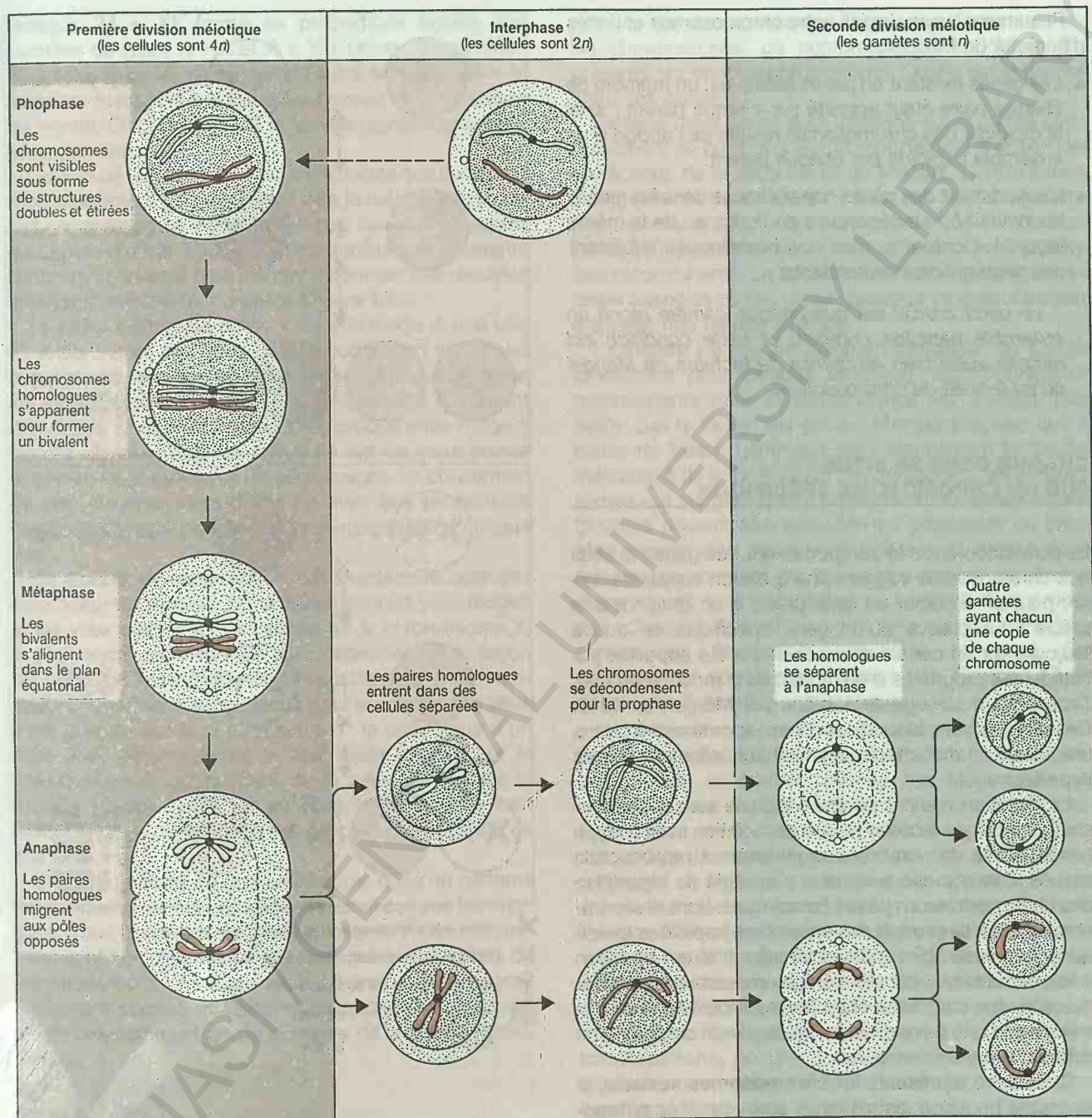


Figure 1.7
La méiose partage en deux le nombre de chromosomes.

La figure montre le comportement d'une paire homologue. Les deux membres de la paire sont dupliqués avant le début de la prophase. Pendant la première division, les homologues s'apparient et ensuite chaque paire de chromatides sœurs migre dans une cellule différente. Pendant la seconde division, les chromatides sœurs sont isolées dans des cellules différentes, et ainsi chaque gamète reçoit une copie.

Résumons les analogies entre chromosomes et unités d'hérédité de Mendel :

- Les gènes existent en paires alléliques, un membre de chaque paire étant apporté par chaque parent ; ainsi la diploïdie des chromosomes résulte de l'apport d'un ensemble haploïde par chaque parent.
- L'assortiment des gènes non alléliques dans les gamètes devrait être indépendant de l'origine ; de la même façon, les chromosomes non homologues subissent une ségrégation indépendante.

Le point crucial est que chaque gamète reçoit un ensemble haploïde complet, et cette condition est remplie aussi bien en termes de facteurs de Mendel qu'en termes de chromosomes.

CHAQUE GÈNE SE SITUE SUR UN CHROMOSOME SPÉCIFIQUE

La corrélation entre le comportement des gènes et celui des chromosomes suggère une question supplémentaire : peut-on associer un gène précis à un chromosome précis ? La preuve qu'un gène spécifique se trouve toujours sur un certain chromosome a été apportée par l'étude des propriétés d'un mutant de la mouche du fruit, *Drosophila melanogaster*, obtenu par Morgan en 1910. Ce mâle aux yeux blancs est apparu spontanément dans une lignée de mouches aux yeux habituellement rouges (type sauvage).

La mutation *blanc* a pu être localisée sur un chromosome spécifique à cause de son association avec un type sexuel. Chez de nombreux organismes à reproduction sexuée, il existe une exception à la règle de répartition des chromosomes en paires homologues dont la séparation à la méiose produit des ensembles haploïdes identiques. Les ensembles mâle et femelle diffèrent de façon évidente dans leur constitution chromosomique, la différence la plus commune étant le remplacement d'un des membres d'une paire d'homologues par un chromosome différent dans un des sexes.

Cette paire représente les **chromosomes sexuels**, et les paires homologues restantes sont appelées **autosomes**. Les effectifs chromosomiques des deux sexes peuvent être désignés $2A + XX$ et $2A + XY$, où A représente l'ensemble haploïde des autosomes et X et Y sont les chromosomes sexuels individuels. Le sexe avec l'effectif homogamétique $2A + XX$ forme uniquement des gamètes du type $A + X$. Le sexe avec l'effectif hétéroga-

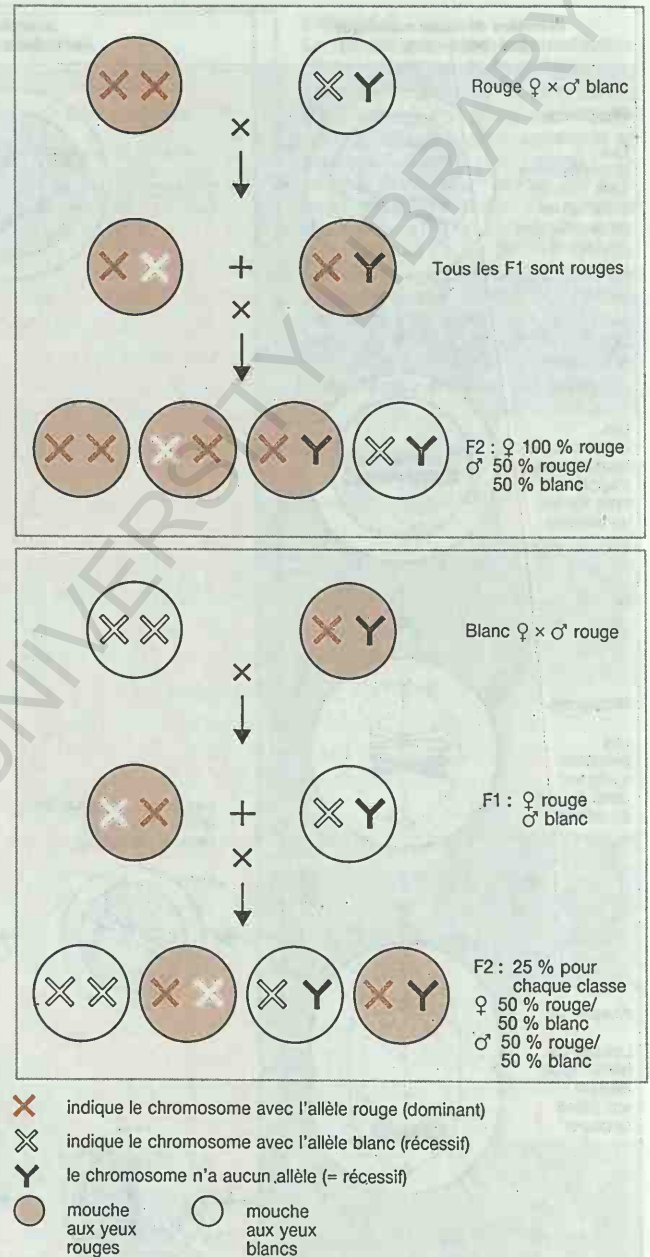


Figure 1.8

Les gènes du chromosome X ont une transmission liée au sexe.

La couleur rouge/blanc des yeux d'une mouche mâle dépend uniquement du chromosome X reçu de sa mère. Le phénotype de la femelle est déterminé selon qu'elle reçoit un allèle rouge dominant de l'un ou l'autre parent. Ceci crée le modèle caractéristique « entrecroisé » de la transmission liée au sexe.

métique $2A + XY$ forme en proportions égales des gamètes de types $A + X$ et $A + Y$. L'union des gamètes d'un sexe avec les gamètes de l'autre sexe perpétue le maintien du rapport des sexes au moment de la formation du zygote. Chez *Drosophila*, la femelle représente le sexe homogamétique.

Les lois de Mendel prévoient entre autres que le résultat d'un croisement génétique doit être le même quelle que soit l'orientation — c'est-à-dire quel que soit le parent qui apporte l'allèle. Mais chez *Drosophila* les croisements réciproques avec le type œil blanc donnent des résultats différents, comme le montre la **Figure 1.8**.

Le croisement mâle blanc $\times$ femelle rouge donne une F1 entière avec des yeux rouges comme on s'y attend si rouge est dominant et blanc est récessif. Mais toutes les mouches aux yeux blancs apparaissant à la F2 sont des mâles. Dans le croisement réciproque mâle rouge $\times$ femelle blanche, tous les mâles F1 ont les yeux blancs et toutes les femelles ont les yeux rouges. Le croisement de ces mouches donne une F2 avec des proportions identiques d'yeux blancs et d'yeux rouges dans chaque sexe.

Ce modèle de transmission suit exactement celui des chromosomes sexuels. Si les allèles pour les yeux rouges et les yeux blancs sont localisés sur le chromosome X, le phénotype de la femelle sera déterminé de la façon habituelle par les allèles présents sur les deux chromosomes X. Toutefois si aucun locus pour la couleur des yeux n'est présent sur le chromosome Y, le phénotype d'un mâle sera déterminé par le seul allèle présent sur le chromosome X, qui provient de la mère. Ainsi dans le modèle typique de **liaison au sexe**, les caractères sont transmis par la mère à tous ses garçons et à chacune de ses filles.

Un gène présent en une seule copie dans un génome qui habituellement est diploïde est décrit comme **hémizygote**. Tous les gènes sur le chromosome Y des mammifères sont hémizygotes ; il en est de même des gènes du chromosome X chez un mâle. L'hémizygotie apparaît également comme une aberration quand (par exemple) un chromosome est perdu au cours de la division cellulaire.

LES GÈNES SONT ALIGNÉS LINÉAIREMENT

Le réassortiment indépendant des chromosomes à la méiose explique le réassortiment indépendant des gènes localisés sur des chromosomes différents. Mais le nombre des facteurs génétiques est beaucoup plus grand que le

nombre des chromosomes. Si tous les gènes sont sur les chromosomes, de nombreux gènes doivent être présents sur chaque chromosome. Quelle est la relation entre ces gènes ?

Lorsque deux gènes sont localisés sur le même chromosome il se peut que leur comportement s'écarte des prédictions de la seconde loi de Mendel. Contrairement aux prévisions de la Figure 1.4, la proportion des génotypes parentaux à la F2 est plus grande que celle prévue parce qu'il y a une *diminution de la formation de génotypes recombinants*. La tendance de certains caractères à rester associés au lieu de se réassortir indépendamment s'appelle une **liaison** (linkage).

La **Figure 1.9** montre comment on utilise un test de croisement pour mesurer la liaison : la proportion de recombinants dans la descendance est d'autant plus petite que la liaison est étroite. Morgan proposa que la cause de liaison génétique est « simplement le résultat mécanique de la localisation des gènes dans les chromosomes ». Il suggéra que la production de classes recombinantes pouvait être assimilée au processus de **crossing-over** visible pendant la méiose. Au début de la méiose, au moment où les quatre copies de chaque chromosome sont organisées en bivalent, il se produit des échanges de matériel (entre deux paires) entre deux chromatides étroitement associées (synapse). Cet échange est appelé un **chiasma** ; la **Figure 1.10** en donne une illustration schématique.

Un chiasma représente un site où deux des chromatides d'un bivalent ont été cassées en un point correspondant — les extrémités cassées ont été croisées et resoudées pour donner de nouvelles chromatides ; chacune d'elles est constituée de matériel provenant d'une chromatide d'un côté du point de jonction et de matériel provenant de l'autre chromatide sur le côté opposé. Les deux chromatides recombinantes ont des structures inverses. C'est un mécanisme de **rupture et de réunion**.

Pour avoir la preuve formelle que le crossing-over est responsable des recombinaisons, il est nécessaire d'établir une corrélation entre la structure chromosomique et le génotype. Ceci est possible grâce à l'existence des **translocations**, où une partie d'un chromosome s'est détachée puis rattachée à un autre. Le chromosome transloqué peut alors être distingué du chromosome normal par son apparence. Par des croisements appropriés, comme ceux de la **Figure 1.11**, il est possible de montrer que la formation de recombinants génétiques se produit uniquement quand il y a eu un crossing-over physique entre les régions chromosomiques appropriées.

Si la probabilité qu'un chiasma se forme entre deux

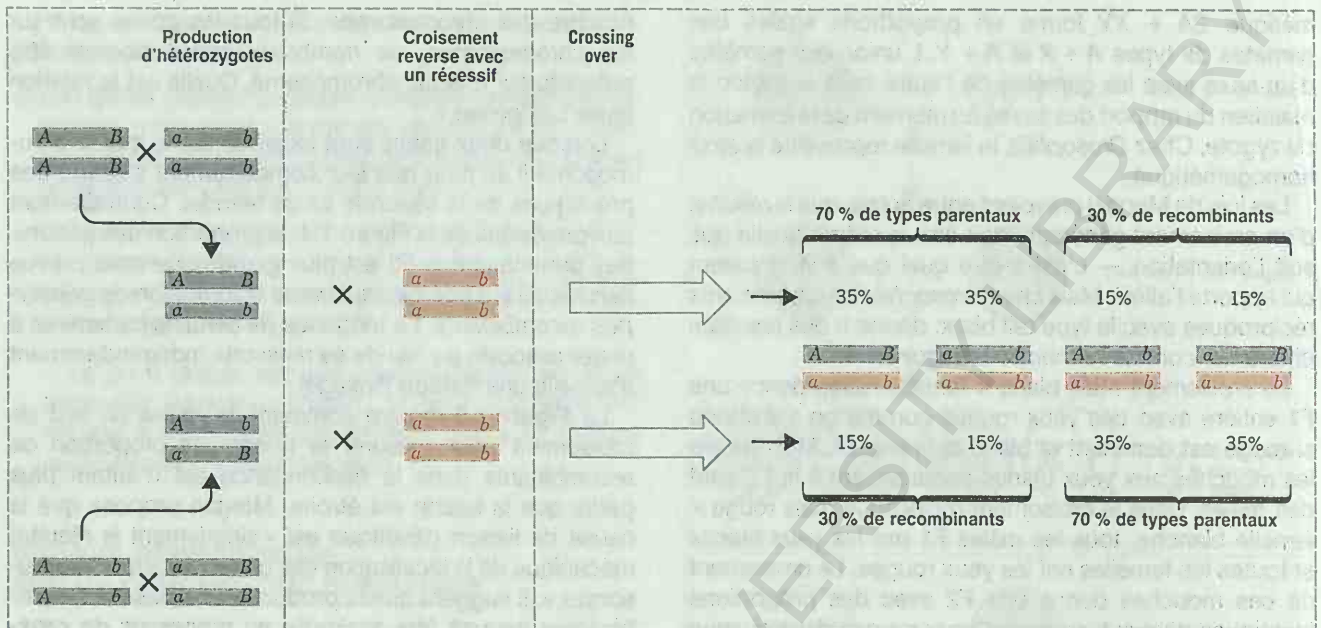


Figure 1.9

La liaison peut être mesurée par un croisement reverse avec un double homozygote.

Les chromosomes sont représentés par des barres ombrées. Le noir représente le matériel provenant des hétérozygotes, le rouge représente le matériel provenant du double homozygote.

Dans le croisement supérieur, un des chromosomes de l'hétérozygote porte les deux allèles dominants, l'autre porte les deux allèles récessifs. Ainsi les types parentaux sont AB et ab. Les types recombinants sont Ab et aB.

Dans le croisement inférieur, un des chromosomes de l'hétérozygote porte un allèle dominant et un récessif ; l'autre chromosome porte la combinaison inverse. Donc Ab et aB sont les types parentaux ; les recombinants sont AB et ab.

Dans chaque croisement la descendance contient une augmentation de la proportion des types parentaux (70 %) et une diminution de la proportion des types recombinants (30 %) comparativement aux 50 % de chaque type attendu avec un réassortiment indépendant. Notons que les deux types parentaux sont présents en quantité égale, de même que les deux types recombinants. La liaison entre A et B est mesurée comme étant 30 % ou 30 unités de recombinaison.

points d'un chromosome dépend de leur distance ; les gènes localisés l'un près de l'autre tendront à rester ensemble. Si la distance augmente, la probabilité de crossing-over entre eux augmentera. Ainsi, si le crossing-over est responsable de la recombinaison, les gènes seront d'autant plus étroitement liés qu'ils seront plus proches l'un de l'autre. Si on inverse le raisonnement, la liaison génétique peut être considérée comme une mesure de la distance physique :

La fréquence de recombinaison entre deux gènes situés sur le même chromosome peut être utilisée comme une **distance sur la carte** pour mesurer leurs localisations relatives. La formule pour mesurer une distance génétique est :

$$\text{Distance sur carte} = \frac{\text{Nombre de recombinants} \times 100}{\text{Nombre total de descendants}}$$

L'**unité génétique** (ou Centi Morgan) est définie comme étant égale à 1 % de crossover. Pour des distances courtes (inférieures à 10 %), le pourcentage de crossover est le même que le pourcentage de recombinants. Cependant, quand deux crossovers apparaissent l'un près de l'autre, ils peuvent rétablir l'arrangement parental des loci sur les deux côtés. Ainsi pour des distances plus longues, l'apparition de doubles crossovers rend le pourcentage de recombinants inférieur au pourcentage de crossovers, et ceci nécessite une correction de la fréquence mesurée.

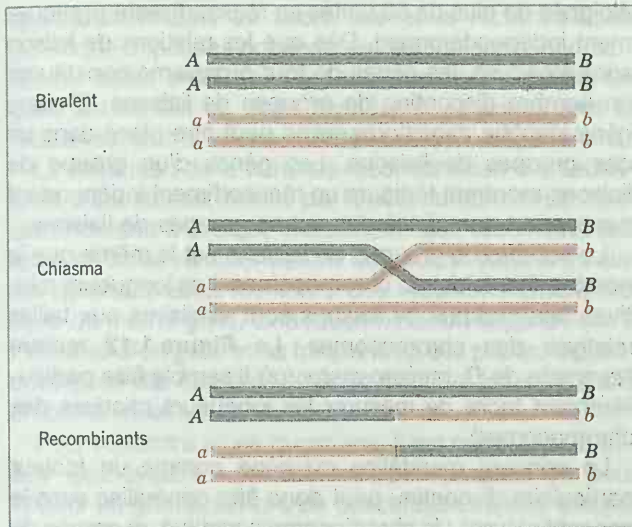


Figure 1.10

La formation d'un chiasma est responsable de l'apparition de recombinants.

Chaque barre représente une chromatide ; les deux chromatides issues d'un parent sont grises, celles de l'autre parent sont rouges. Elles s'associent pour former un bivalent à la prophase méiotique. (Puisque le bivalent contient quatre chromatides, ceci est quelquefois rapporté comme étant le « stade à quatre brins » de la méiose).

Un **crossing-over** se produit entre deux des chromatides, quand après cassure de celles-ci au même site, les extrémités opposées sont reliées. Ceci crée les deux chromatides recombinants (qui portent les

nouvelles combinaisons Ab et aB). Les deux autres chromatides restent de type parental (AB et ab).

La représentation du crossing-over est purement schématique et n'implique pas qu'il y ait en fait un changement dans l'appariement des partenaires sur toute la longueur du chromosome (en fait les chromatides sœurs restent appariées après la formation du chiasma, événement local impliquant un changement à un site spécifique).

Formation du chiasma et cartographie génétique

Les caractéristiques de la formation du chiasma découlent de la recombinaison génétique. L'apparition d'un chiasma réduit la probabilité qu'un autre apparaisse dans le voisinage. Cet effet est appelé **interférence positive**. Il réduit le nombre de doubles recombinants dans un test de croisement à trois points.

Si on considère le croisement en termes de gamètes, la combinaison $ABC \times abc$ donne les **doubles recombinants** AbC et aBc quand des cross-overs se produisent entre A et B et aussi entre B et C sur le même chromosome. Si les deux cross-overs sont indépendants, la fréquence de doubles cross-overs sera le produit des fréquences des événements séparés. Cependant à cause de l'interférence positive, la fréquence réelle est inférieure à celle prévue par ce calcul.

Les fréquences de recombinaison mesurées doivent être corrigées pour compenser l'apparition des doubles cross-overs. Les doubles cross-overs réduisent la fréquence de recombinaison, ainsi une fréquence mesurée à 30 % correspondra à > à 30 unités de recombinaison.

Quand des cross-overs adjacents se produisent, le fait qu'une paire des quatre chromatides soit impliquée dans un cross-over est indépendant du fait qu'une paire soit impliquée dans un autre cross-over.

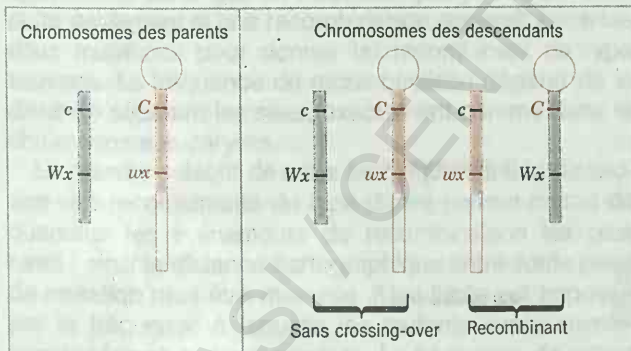


Figure 1.11

La recombinaison génétique résulte d'un crossing-over.

Chez un mutant du maïs le chromosome qui porte les allèles C et wx possède un « bouton » à une extrémité, alors que l'autre extrémité a une longue translocation de matériel provenant d'un autre chromosome. Chez un hétérozygote possédant un chromosome normal porteur des allèles c et Wx , la recombinaison entre des marqueurs génétiques est toujours associée à la formation d'un nouveau type de chromosome.

On peut faire une observation importante quand des caractères multiples sont suivis en même temps. Pour les gènes portés par un même chromosome, les *distances génétiques individuelles* sont (approximativement) additives. Ainsi, si deux gènes A et B sont distants de 10 unités, et si un gène C se trouve 5 unités plus loin que B , la mesure directe de la distance séparant A et C sera proche de 15. Les gènes peuvent alors être placés dans un **ordre linéaire**.

Pour la construction d'une carte génétique, il faut savoir que la distance entre les gènes ne dépend pas des *allèles* spécifiques qui sont utilisés, mais ne dépend que des **loci** des gènes. Le locus définit la *position* occupée sur le chromosome par le gène représentant un caractère particulier. Les *diverses formes possibles du gène* — c'est-à-dire les allèles utilisés dans la cartographie — *ont toutes la même localisation*.

Ainsi la cartographie génétique a pour but d'identifier les positions des loci des gènes, qui sont fixes et disposés dans un ordre linéaire. Dans une expérience de cartogra-

phie, le même résultat est obtenu quelle que soit la combinaison particulière des allèles (voir Figure 1.9).

La liaison ne se manifeste pas entre toutes les paires de gènes localisées sur un même chromosome. La recombinaison maximum entre deux loci est de 50 %, correspondant à la ségrégation indépendante prédite par la seconde loi de Mendel. (Bien qu'il y ait une forte probabilité pour qu'une recombinaison apparaisse entre deux gènes situés loin l'un de l'autre sur un chromosome, chaque recombinaison individuelle implique seulement deux des quatre chromatides associées, et ainsi il y a une limite de recombinaison égale à 50 % entre les gènes).

Malgré leur présence sur le même chromosome, les gènes qui sont bien séparés peuvent se réassortir indépendamment. Cependant, bien qu'ils ne montrent aucune liaison directe, chacun peut être lié à des gènes situés entre eux, et de cette façon une carte génétique peut être étendue au-delà de la limite des 50 % de recombinaison directement mesurable entre toute paire de gènes. Une carte génétique est habituellement basée sur des mesures impliquant des gènes assez proches, et peut être sujette à des corrections lorsqu'elle est établie à partir du simple pourcentage de recombinaison.

Un **groupe de liaison** inclut tous les gènes qui peuvent être reliés soit directement, soit indirectement par des relations de liaison. Les gènes qui sont proches les uns des autres montrent une liaison directe ; ceux qui sont

éloignés de plus de 50 unités se réassortissent pratiquement indépendamment. Dès que les relations de liaison sont étendues, les gènes de tout organisme constituent un nombre discontinu de groupes de liaisons. Chaque gène identifié dans l'organisme peut être placé dans un des groupes de liaisons. Les gènes d'un groupe de liaisons montrent toujours un réassortiment indépendant par rapport aux gènes des autres groupes de liaisons.

Le nombre de groupes de liaisons est le même que le nombre (haploïde) de chromosomes. Les longueurs relatives des groupes de liaisons sont similaires aux tailles relatives des chromosomes. La **Figure 1.12** montre l'exemple de *D. melanogaster* (où il semble être particulièrement facile de mesurer les longueurs relatives des chromosomes).

Le concept mendélien du gène comme un facteur particulière discontinu peut donc être généralisé avec le concept suivant : le chromosome constitue un groupe de liaisons, divisé en de nombreux gènes dont l'arrangement physique peut être à la base de leur comportement génétique.

LA CARTE GÉNÉTIQUE EST CONTINUE

Sur les cartes génétiques des organismes supérieurs établies pendant la première moitié de ce siècle, les gènes sont disposés comme des perles sur un collier. Ils

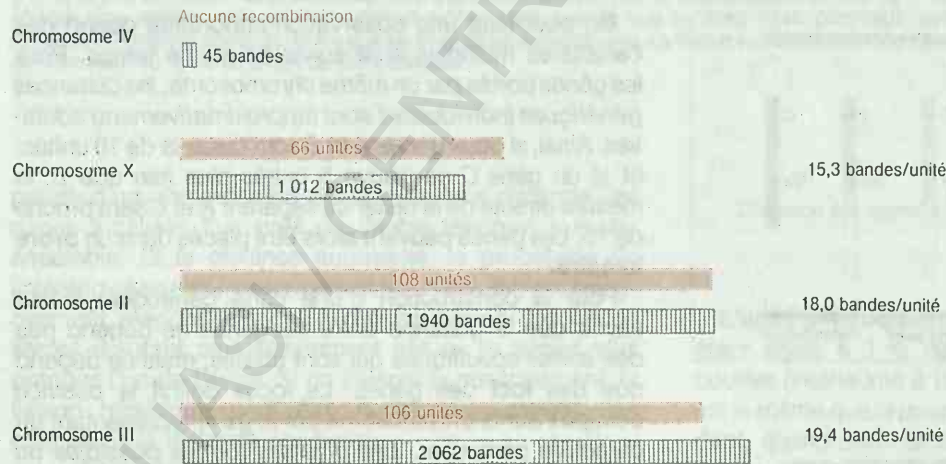


Figure 1.12

La carte de liaison peut être rapportée à la taille physique des chromosomes.

Chaque chromosome de *D. melanogaster* peut être représenté comme un nombre de « bandes ». Le nombre relatif de bandes reflète la longueur physique de chaque chromosome. Le nombre relatif d'unités de recombinaison par chromosome est identique.

apparaissent dans un ordre fixe et la recombinaison génétique implique l'échange de régions correspondantes du collier entre des chromosomes homologues. Le gène est en fait un objet mystérieux (la perle) dont la relation avec un environnement (le collier) n'est pas claire.

La résolution de la carte de recombinaison d'un eucaryote supérieur est réduite par le petit nombre de descendants obtenu à chaque croisement. Une recombinaison apparaît aussi rarement entre des points voisins qu'elle est rarement observée entre des mutations différentes dans un même gène. Une recombinaison apparaît-elle à l'intérieur d'un gène et sa fréquence à ces postes rapprochés peut-elle être utilisée pour disposer les sites dans un ordre linéaire ?

Pour répondre à ces questions, il est nécessaire d'utiliser des systèmes microbiens avec lesquels on peut obtenir un grand nombre de descendants après chaque croisement génétique. Un système convenable est fourni par le **phage T4**, un virus qui infecte la bactérie *Escherichia coli*. L'infection d'une seule bactérie permet la production d'environ 100 phages en moins de 30 minutes. Les phages sont libérés quand la bactérie est lysée (elle s'ouvre tout d'un coup).

La constitution d'un locus particulier fut étudiée intensivement par Benzer avec les gènes *rII* du phage ; ces gènes sont responsables d'un changement dans la façon de tuer les bactéries, connu sous le nom de **lyse rapide**. Quand deux phages différents, mutants *rII*, sont utilisés pour infecter une bactérie, les conditions peuvent être ajustées de sorte que de nouveaux phages seront produits *seulement* si une recombinaison apparaît entre les deux mutations pour donner un recombinant de type sauvage. La fréquence de recombinaison dépend de la distance séparant les sites, exactement comme dans le chromosome eucaryote.

Le pouvoir sélectif de cette technique dans la distinction des recombinants du type désiré permet même de quantifier les événements de recombinaison les plus rares ; ainsi la distance cartographique entre *toute* paire de mutation peut être mesurée. (Une limite est imposée par la fréquence à laquelle les mutants « retournent » spontanément au type sauvage. La fréquence de retour est d'environ 0,0001 %, donc pour atteindre un niveau minimum de mesure 10 fois plus élevé que ce « bruit de fond », seules les fréquences de recombinaison supérieures à 0,001 % peuvent être mesurées.)

Environ 2 400 mutations apparaissent dans 304 sites mutants différents. (Quand deux mutations ne peuvent pas recombiner, on suppose qu'elles sont apparues de façon indépendante et spontanée au *même* site généti-

que.) Les mutations ont pu être disposées de façon linéaire, montrant que le gène lui-même a une construction linéaire identique à l'alignement des gènes sur un chromosome. Ainsi la carte génétique est linéaire aussi bien à l'intérieur qu'entre les loci ; c'est-à-dire, qu'elle représente une séquence continue à l'intérieur de laquelle résident les gènes.

UN GÈNE - UNE PROTÉINE

Jusqu'en 1945, le gène était considéré comme l'unité fondamentale d'hérédité, mais il n'y avait aucune explication universelle à sa fonction. Les gènes pouvaient être identifiés uniquement par les mutations qui entraînaient une aberration dans le phénotype. La principale difficulté était de pouvoir compter sur ces mutants qui étaient disponibles mais qui n'étaient pas nécessairement appropriés pour des études biochimiques. Ainsi s'accumulaient de nombreux exemples dans lesquels des anomalies de réactions biochimiques précises pouvaient être associées à des mutations spécifiques, mais le rôle du gène restait inconnu.

Un essai systématique pour associer gènes et enzymes fut démarré en 1930 par Beadle et Ephrussi, dès qu'ils eurent conclu que le développement de la couleur oculaire normale (rouge) de *Drosophila* passe par une série d'étapes. Un blocage à différentes étapes entraîne la production de couleurs mutantes différentes. Mais il fallut attendre très longtemps pour que la voie complète pût être résolue. Beadle et Tatum essayèrent d'aborder le problème dans l'autre direction. Ainsi Beadle se souvenait : « il m'apparut subitement qu'il devait être possible d'inverser la procédure que nous avions suivie et qu'au lieu d'essayer de résoudre la chimie des différences génétiques connues, nous devrions être capables de sélectionner des mutants chez lesquels des réactions chimiques connues étaient bloquées ».

Avec le champignon *Neurospora crassa*, des mutants furent créés par irradiation aux rayons X et sélectionnés pour leur incapacité à croître sur un milieu minimum. La nature biochimique du défaut a pu être identifiée en découvrant le facteur de croissance qui permettait à la souche mutée de croître. Chaque mutant se trouvait être bloqué à une étape métabolique précise, réalisée dans la souche de type sauvage par une seule enzyme. Le blocage à chaque étape entraîne une accumulation de l'intermédiaire métabolique immédiatement antérieur à l'étape.

La **Figure 1.13** illustre la voie reconnue par la suite

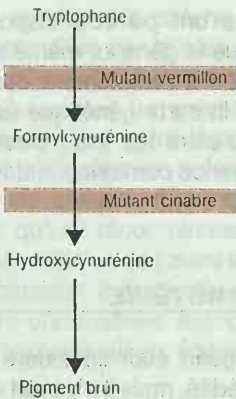


Figure 1.13
Les gènes contrôlent les étapes métaboliques.

Certaines mutations qui affectent la couleur oculaire chez *D. melanogaster*, agissent en bloquant différentes étapes de la voie de transformation du tryptophane en pigment brun. Chaque mutation provoque l'absence d'une enzyme particulière, conduisant à l'accumulation de l'intermédiaire sur lequel il agit. La couleur de l'œil est influencée par l'effet (ou l'absence d'effet) de cet intermédiaire.

comme étant responsable de la production du pigment oculaire brun chez *D. melanogaster* ; un blocage à différentes étapes de cette voie provoque différents phénotypes qui dépendent de l'intermédiaire métabolique accumulé.

A partir de 1945 les résultats de l'analyse ont commencé à être vulgarisés sous la forme de l'**hypothèse un gène : une enzyme**. Ceci supposait que chaque

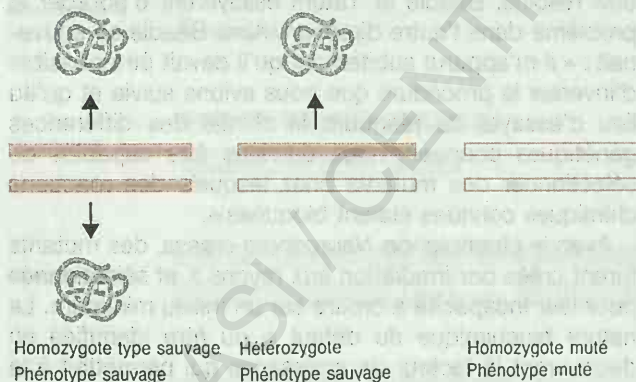


Figure 1.14
Les gènes codent pour des protéines.

Chez le type sauvage, les deux allèles (barres pleines) sont actifs et produisent une protéine. Chez l'hétérozygote, l'allèle dominant (barre pleine) est actif et produit une protéine, mais l'allèle récessif (barre vide) ne produit aucune protéine active. Chez un homozygote possédant deux allèles récessifs aucune protéine n'est produite ; ainsi l'organisme perd cette fonction.

étape métabolique est catalysée par une enzyme spécifique dont la production est sous la responsabilité d'un seul gène. Une mutation dans le gène peut altérer l'activité de la protéine dont il est responsable.

Puisqu'une mutation est un événement aléatoire vis-à-vis de la structure du gène, il est donc hautement probable qu'elle endommagera la fonction du gène. Ainsi la majorité des mutations produit un gène non fonctionnel ; mais ceci n'est en aucune façon la seule conséquence de la mutation, qui modifie quelquefois la fonction du gène plus qu'elle ne l'abolit.

Ces notions expliquent la nature des mutations récessives : elles représentent une absence de fonction, parce que le gène muté n'a pas pu produire son enzyme habituelle. Comme décrit dans la **Figure 1.14**, cependant, chez un hétérozygote contenant un allèle de type sauvage et un allèle muté, l'allèle de type sauvage est capable de diriger la production de l'enzyme. L'allèle de type sauvage est alors dominant.

(Ceci suppose qu'une *quantité* adéquate de protéine est produite par le seul allèle de type sauvage. Quand ceci n'est pas vérifié, la plus petite quantité de protéine fabriquée par un seul allèle (comparativement à celle fabriquée par deux) peut produire le phénotype intermédiaire d'un allèle partiellement dominant chez un hétérozygote.)

On dut attendre jusqu'en 1957 pour avoir la preuve directe qu'un gène est réellement responsable du contrôle de la structure d'une protéine. En effet, Ingram montra que le caractère gène unique de l'anémie falciforme pouvait être expliqué par un changement dans la composition en acides aminés de l'hémoglobine.

Une modification de l'hypothèse est nécessaire pour les protéines constituées de plus d'une sous-unité. Si les sous-unités sont toutes identiques, la protéine est un **homooligomère**, représentée par un seul gène. Si les sous-unités sont différentes la protéine est un **hétéroooligomère**.

L'hémoglobine est un exemple de protéine constituée de plusieurs types de chaînes polypeptidiques ; un groupe hème (qui fixe le fer) est associé à deux sous-unités α et à deux sous-unités β . Chaque type de sous-unité comporte une chaîne polypeptidique différente codée par son propre gène. Ainsi la fonction de l'hémoglobine peut être altérée par une mutation dans les gènes codant pour l'un ou l'autre des polypeptides α ou β .

Énoncée sous une forme plus générale, applicable à toute protéine hétéroooligomérique, l'hypothèse un gène : une enzyme sera exprimée plus précisément sous la forme **un gène : une chaîne polypeptidique**.

UNE DÉFINITION DÉTAILLÉE : LE CISTRON

Si une mutation récessive est produite par tout changement dans un gène qui empêche la production d'une protéine active, il devrait y avoir un grand nombre de telles mutations dans n'importe quel gène ; de nombreux remplacements d'acides aminés doivent pouvoir suffisamment changer la structure d'une protéine pour entraîner sa fonction. Les différents variants d'un même gène sont appelés des **allèles multiples**, et leur existence rend possible l'apparition d'un hétérozygote entre allèles mutants.

Quand deux mutations ont le même effet phénotypique et se localisent proches l'une de l'autre, elles peuvent intéresser des allèles. Toutefois, elles pourraient également représenter des mutations dans *deux* gènes *différents* dont les protéines sont impliquées dans la même fonction. Le **test de complémentation** est utilisé pour déterminer si deux mutations résident dans le même gène ou dans des gènes différents. Le test consiste à fabriquer un hétérozygote pour les deux mutations (en accouplant des parents homozygotes pour chaque mutation).

Si les mutations résident dans le même gène, les génotypes parentaux peuvent être représentés ainsi

$$\frac{m_1}{m_1} \text{ et } \frac{m_2}{m_2}$$

Le premier parent fournit un allèle muté m_1 et le second parent fournit un allèle muté m_2 ; ainsi l'hétérozygote a la composition

$$\frac{m_1}{m_2}$$

dans laquelle *aucun gène de type sauvage n'est présent*, l'hétérozygote a donc un phénotype mutant.

Si les mutations résident dans des gènes différents, les génotypes parentaux peuvent être représentés ainsi

$$\frac{m_1 +}{m_1 +} \text{ et } \frac{+ m_2}{+ m_2}$$

où chaque chromosome a une copie de type sauvage d'un gène (représentée par le signe plus) et une copie mutée de l'autre. L'hétérozygote a donc la constitution

$$\frac{m_1 +}{+ m_2}$$

dans laquelle les deux parents ont apporté une copie de type sauvage de chaque gène. L'hétérozygote a le type sauvage ; on dit que les deux gènes ont **complémenté**.

La **Figure 1.15** fournit une description plus élaborée du test de complémentation. Si nous considérons uniquement les sites individuels de mutation (sans tenir compte s'ils résident dans le même gène ou dans des gènes différents), le double hétérozygote peut avoir l'une ou l'autre des deux configurations. Dans la configuration *cis*, les deux mutations sont présentes sur le même chromosome. Dans la configuration *trans*, elles sont présentes sur les chromosomes opposés. Les effets relatifs de ces configurations sont déterminés selon que les mutations résident dans le même gène ou dans des gènes différents.

Considérons d'abord la situation où les mutations sont dans le même gène. La configuration *trans* correspond au test que nous venons de décrire. Les deux copies du gène sont mutées. Dans la configuration *cis*, cependant, un des génomes élabore une protéine qui a deux mutations, alors que l'autre n'a rien et est donc de type sauvage. Ainsi quand deux mutations sont dans le même gène, le phénotype d'un hétérozygote est déterminé par la configuration *cis* ou *trans*. Il est mutant quand les mutations sont en *trans* et de type sauvage quand elles sont en *cis*.

Au contraire, quand les mutations sont localisées dans des gènes différents, la configuration n'a aucun effet. Dans chaque cas, il y a une copie de chaque gène muté et une copie de chaque gène de type sauvage.

Ainsi lorsque deux mutations sont testées en configuration *cis*, on obtient le même résultat qu'elles soient localisées dans le même gène ou dans des gènes différents. Ainsi la configuration *cis* est utilisée comme contrôle pour mettre en évidence la présence d'une fonction de type sauvage.

La complémentation est testée en pratique en déterminant si l'hétérozygote *trans* montre un phénotype sauvage (les mutations sont dans des gènes différents) ou un phénotype muté (les mutations sont dans le même gène). Pour les eucaryotes, l'expérience est effectuée en construisant le double hétérozygote approprié. Pour les virus, une cellule hôte est simultanément infectée avec les deux types mutés.

Quand deux mutations *ne peuvent* complémenter en *trans*, on en conclut que les deux affectent la même fonction. Elles appartiennent alors au même **groupe de complémentation**. Dans l'analyse de la région *rII* du phage T4, les mutations constituent deux groupes de complémentation, chaque groupe constituant un groupe contigu aux sites mutés. Aucune paire de mutations du groupe *rIIA* ne pouvait complémenter, aucune paire de mutations du groupe *rIIB* ne pouvait complémenter ; mais n'importe quelle mutation du groupe *rIIA* pouvait complé-

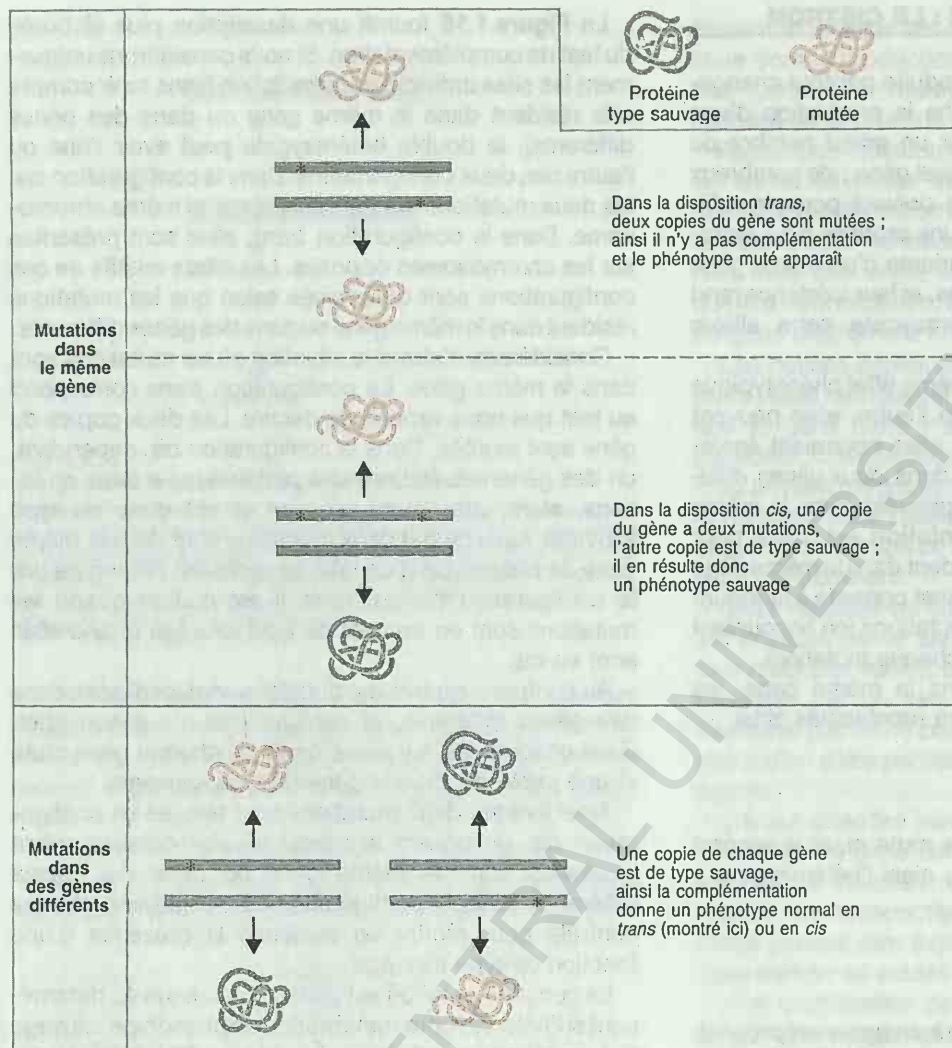


Figure 1.15

Le cistron est défini par le test de complémentation.

Les gènes sont représentés par des barres ; les astérisques identifient les sites de mutations.

menter avec n'importe quelle mutation de *rII*B. Ainsi les mutations *rII* identifient deux groupes de complémentation adjacents, chacun avec la même fonction phénotypique.

L'unité génétique définie par le test de complémentation est formellement appelée le **cistron**. Deux mutations dans le même cistron ne peuvent pas compléter en *trans* ; l'apparition de complémentation montre que les mutations résident dans des cistrons *différents*. Le terme « cistron » est utilisé de façon interchangeable avec l'ancien terme « gène » et d'une façon générale est en train de tomber en désuétude. Dans le système *rII*,

chaque groupe de complémentation identifie un cistron ou un gène différent.

Si deux gènes complètent en *trans*, chacun doit être responsable de la production d'une protéine qui est capable de coopérer avec la protéine produite par l'autre gène, de façon à créer ensemble un phénotype sauvage. On dit que les produits sont *trans-agissants* et on conclut qu'ils représentent des molécules diffusibles capables d'agir ensemble indépendamment de leurs origines génomiques.

Une exception à la règle que seuls des gènes différents peuvent compléter est quelquefois rencontrée quand

un gène représente un polypeptide qui est la sous-unité d'une protéine homooligomérique. Dans la cellule de type sauvage, la protéine est constituée de plusieurs sous-unités *identiques*. Dans une cellule contenant deux allèles mutés, cependant, leurs produits peuvent se mélanger pour former des protéines oligomériques qui contiennent les *deux types* de sous-unités. Parfois les deux mutations se compensent, de sorte que la protéine à sous-unités mixtes est active, même si les protéines constituées de chaque type de sous-unités mutées sont inactives. Cet effet est appelé **complémentation interallélique**.

CARTOGRAPHIE DES MUTATIONS AU NIVEAU MOLÉCULAIRE

L'isolement des **revertants** est une caractéristique importante qui distingue les mutations ponctuelles et les insertions des délétions. Une **mutation ponctuelle** est un changement qui affecte une seule paire de bases dans un gène (telle que la substitution d'une paire de bases par une autre). Elle peut se reverser par restauration de la séquence originale ou par une mutation compensatrice ailleurs dans le gène. Une **insertion** de matériel supplémentaire peut reverser par délétion du matériel inséré. (Des mutations peuvent également apparaître dans d'autres gènes pour circonvier les effets d'une mutation dans le gène d'origine. Cet effet est appelé **suppression**, ou, plus formellement, suppression interchromosomique.)

La **délétion** d'une partie d'un gène ne peut pas reverser. De telles délétions ont une grande utilité dans la cartographie génétique. L'étendue génétique d'une délétion est définie par son incapacité à recombiner avec une série de mutations adjacentes (parce qu'il manque la partie correspondante de l'ADN de type sauvage). La délétion peut recombiner avec des mutations ponctuelles situées de part et d'autre. La **Figure 1.16** montre qu'on peut visualiser génétiquement une délétion grâce aux transitions entre recombinaison et non recombinaison avec une série de mutations ponctuelles.

Avec une série de délétions partiellement chevauchantes, on peut cartographier toute mutation ponctuelle en testant sa capacité à recombiner avec celles-ci. La **Figure 1.17** illustre le protocole. Quand deux délétions ne peuvent pas recombiner avec une mutation ponctuelle, le site de mutation doit être dans la région commune aux délétions. Quand une délétion recombine et l'autre pas, le site de mutation doit se trouver dans la région où les mutations ne se recouvrent pas.

Pour caractériser un génome, il serait idéal de pouvoir

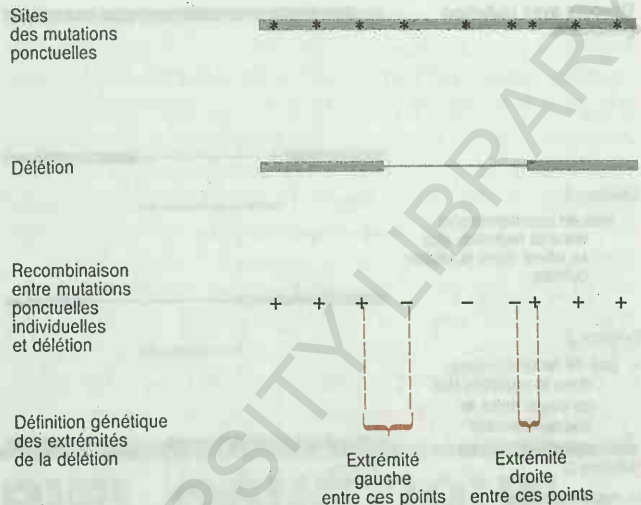


Figure 1.16

Une délétion se recombina avec une série de mutations ponctuelles située de chaque côté de la région déléetée, mais ne peut pas recombiner avec des mutations ponctuelles situées à l'intérieur de la région déléetée. Les limites de la délétion sont indiquées par le passage entre recombinaison et non recombinaison.

muter chaque gène de l'organisme. Mais l'application de cette approche est doublement restreinte.

D'abord, il est nécessaire d'avoir un critère de distinction entre le mutant et le type sauvage. Le critère le plus simple est un changement visible, par exemple, de la couleur de l'œil. D'autres critères peuvent être établis en imaginant des procédures sélectives convenables ; par exemple, en ajustant les conditions de croissance de sorte que l'absence ou la présence d'une enzyme soit requise pour la survie.

Le problème soulevé par cette approche est qu'il peut exister des fonctions dont nous ignorons l'existence et pour lesquelles nous ne pourrions donc pas imaginer de tests convenables. A l'extrême limite on peut aussi imaginer l'existence de gènes dont les produits sont inutiles et dont l'absence n'a alors aucun effet (au moins sous les conditions que nous utilisons). Comment les détecter ?

La seconde difficulté concerne les fonctions *essentielles* pour la viabilité. Une mutation dans un tel gène doit vraisemblablement tuer l'organisme. Pour isoler les mutations dans ces gènes, elles doivent apparaître sous la forme de mutations **létales conditionnelles**. Ces mutations sont létales sous un ensemble de conditions, mais ne provoquent aucun défaut, ou seulement un défaut très réduit, sous des conditions appropriées qui permettent à la cellule de se perpétuer.

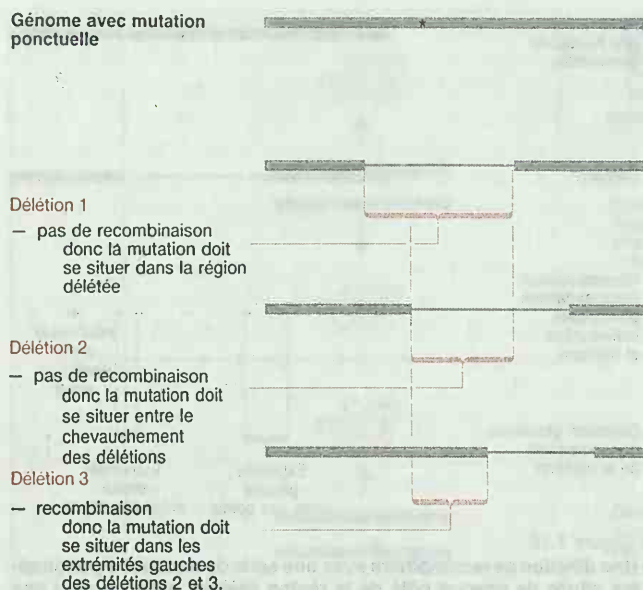


Figure 1.17

La cartographie des délétions peut être utilisée pour localiser des mutations ponctuelles entre les extrémités de délétions chevauchantes.

Si une mutation ponctuelle ne peut pas recombiner avec une délétion, le site de mutation peut s'étendre à l'intérieur d'une région qui a été délétée. Si une mutation ponctuelle recombine avec une délétion, le site de la mutation peut s'étendre en dehors de la région délétée. En comparant la capacité d'une série de délétions à recombiner avec une mutation ponctuelle, le site de la mutation peut être identifié.

Ainsi le même organisme (muté) peut être étudié dans deux conditions. Dans des conditions **permissives**, il ne manifeste pas le phénotype muté et peut alors être perpétué. Dans des conditions **non permissives**, il meurt ou devient sévèrement malade, si bien qu'il peut être étudié pendant la transition entre les conditions permissives et non permissives. (Bien sûr, les mutations conditionnelles ne sont pas restreintes aux fonctions essentielles, mais en principe peuvent être trouvées dans n'importe quel gène.)

La terminologie utilisée pour décrire les mutations est résumée dans la **Note 1.1**.

POUR EN SAVOIR PLUS

La notion moderne de gène/cistron prit naissance avec l'article de **Benzer** (*Proc. Nat. Acad. Sci. USA* 41, 344, 1955). Le travail sur les bactériophages est passé en revue dans un volume édité par **Cairns, Stent et Watson**, *Phage and the Origins of Molecular Biology* (Cold Spring Harbor Laboratory, New York, 1966).

NOTE 1.1

La Terminologie des Mutations

La terminologie utilisée pour décrire les loci génétiques et leurs effets varie quelque peu entre les différents organismes, mais il y a des termes et des caractéristiques communs dont nous aurons besoin.

Marqueur génétique est un terme quelquefois utilisé pour décrire un gène intéressant, par exemple, un gène utilisé dans une expérience de cartographie ou pour identifier une région particulière. Ainsi une cellule peut comporter une classe déterminée de marqueurs, c'est-à-dire des allèles.

Les caractéristiques du génotype sont *toujours écrites en italique* ; le phénotype est décrit par les mêmes termes, mais non exprimé en italique.

Type sauvage peut être utilisé en référence à la forme active habituelle d'un gène ou d'un caractère phénotypique. Dans certains systèmes, il est indiqué par un exposant plus après le nom du locus (W^+ est l'allèle de type sauvage pour la couleur oculaire (rouge) chez *D. melanogaster*). Quelquefois + est lui-même utilisé pour désigner l'allèle de type sauvage, et seuls les allèles mutés sont indiqués par le nom du locus.

Une forme totalement déficitaire du gène (ou une absence d'expression) peut être indiquée par le signe — en exposant. Quand il y a une variété d'allèles mutés avec différents effets, d'autres exposants peuvent être introduits pour une meilleure distinction. Par exemple, il y a de nombreux mutants du locus *W*, incluant W' (couleur oculaire ivoire), W^a (couleur oculaire abricot), et ainsi de suite.

Les gènes reçoivent des abréviations dont la nature dépend du type de l'organisme. Dans la plupart des systèmes eucaryotes, il n'y a aucune restriction sur la forme de l'abréviation. Par convention la première lettre de l'abréviation est en majuscule pour un allèle dominant (chez la levure le nom complet du locus est en majuscules), alors qu'un allèle récessif est entièrement en minuscules.

Chez les bactéries, la nomenclature standard utilise une abréviation à trois lettres minuscules pour tous les gènes qui influent sur une caractéristique spécifique. Chaque gène individuel reçoit une lettre majuscule supplémentaire. Ainsi il existe une série de gènes *lac*, connus comme étant *lac Z*, *lac Y*, et *lac A*. L'allèle type sauvage de *lac Z* serait *lac Z⁺* ; un mutant déficitaire serait *lac Z⁻*. Les mêmes abréviations sont utilisées pour décrire le phénotype, mais la première lettre est une majuscule. Ainsi les allèles *lac* type sauvage ou mutés donneront respectivement naissance aux phénotypes bactériens *Lac⁺* ou *Lac⁻*. Le produit d'un gène est un caractère du phénotype et y est donc rattaché ; par exemple, la protéine *Lac Z* est l'enzyme β -galactosidase.

CHAPITRE 2

L'ADN EST LE MATÉRIEL GÉNÉTIQUE

L'information est transmise sous deux formes d'une génération à l'autre. L'œuf fertilisé (chez les espèces à reproduction sexuée) ou la cellule fille (chez les espèces à reproduction asexuée) reçoit un ensemble de structures organisées préexistantes — son existence même reflète les caractéristiques de la structure cellulaire. De plus il reçoit un ensemble de gènes, nécessaires à la fabrication des structures futures pendant le développement de l'organisme. Chaque gène agit en représentant une chaîne polypeptidique définie, ce qui nous confronte au problème de savoir comment une séquence de nucléotides dans l'ADN représente une séquence d'acides aminés dans une protéine.

Une des caractéristiques fondamentales de la structure de l'ADN est qu'elle est *indépendante de la séquence spécifique de ses nucléotides constitutifs*. La séquence des nucléotides, éléments de construction d'un acide nucléique, est importante, non pas au niveau de la structure elle-même, mais parce qu'elle **code** pour la séquence des acides aminés qui constituent le polypeptide correspondant. Le concept que chaque protéine est constituée d'une série définie d'acides aminés date des années 1950 après que Sanger eût caractérisé l'insuline. La relation entre une séquence d'ADN et la séquence d'une protéine correspondante est appelée le **code génétique**.

La structure et/ou l'activité enzymatique de chaque protéine dépendent de sa séquence primaire en acides aminés. Ainsi en déterminant la séquence des acides aminés dans chaque protéine, le gène est capable de

porter toute l'information nécessaire pour déterminer une chaîne polypeptidique active. De cette façon, un seul type de structure — le gène — peut se présenter sous un grand nombre de formes polypeptidiques.

L'ensemble des divers produits protéiques d'une cellule assume les activités catalytiques et structurales chargées de définir son type. Évidemment, en plus des séquences géniques qui codent pour des protéines, l'ADN contient également certaines séquences dont la fonction est d'être reconnues par une molécule régulatrice, généralement une protéine. Dans ce cas la fonction de l'ADN est directement déterminée par sa séquence, et non via un code intermédiaire. Les deux types de région, les gènes exprimés sous forme de protéines et les séquences reconnues comme telles, constituent l'information génétique.

Les mutations sont des altérations héréditaires qui modifient l'information génétique. Après la découverte de l'ADN comme étant le matériel génétique, le concept qu'une mutation est un changement dans la séquence des nucléotides, vient naturellement. A cause du code génétique, une modification dans la séquence nucléotidique peut entraîner une modification dans la séquence en acides aminés, et ainsi altérer l'activité de la protéine.

DÉCOUVERTE DE L'ADN

Le matériel génétique fonctionne en vertu de sa capacité à déterminer une grande variété de protéines. Les première-

res idées sur la nature du matériel génétique furent faussées par une supposition erronée : la structure du matériel génétique devait être aussi complexe que celle des protéines dont il contrôle la production. En fait, on a pensé pendant longtemps que *seules* les protéines pouvaient avoir assez de diversité pour déterminer d'autres protéines. Cette hypothèse fut rejetée quand on réalisa que le matériel génétique porte l'information nécessaire à la synthèse d'une protéine sous une forme codée.

L'idée que le matériel génétique est un acide nucléique prit forme en 1928 avec la découverte de Griffith sur la **transformation**. La bactérie *Pneumococcus* tue les souris par pneumonie. La virulence de la bactérie est déterminée par la **capsule polysaccharidique**, un composant de la surface cellulaire. Plusieurs **types** (I, II, III) de *Pneumococcus* possèdent différents polysaccharides capsulaires. Tous ces types ont une apparence **lisse** (L), due à la présence du polysaccharide capsulaire à la surface.

Naturellement, les bactéries lisses ne peuvent tuer les souris que lorsqu'elles sont capables de se multiplier. Si les bactéries sont tuées par un traitement thermique, elles perdent leur capacité à nuire à l'animal.

Chacun des types lisses de *Pneumococcus* peut donner des variants incapables de produire le polysaccharide capsulaire. Ces bactéries ont une surface **rugueuse** (R).

Elles sont **non virulentes** ; elles ne tuent pas les souris (car en l'absence du polysaccharide, les bactéries sont soumises à des attaques puis à une destruction dans l'animal ; le manteau polysaccharidique est nécessaire à la protection de la bactérie).

L'expérience illustrée dans la **Figure 2.1** montre que les bactéries L inactivées thermiquement et les bactéries R non virulentes mises ensemble ont un effet assez différent de celui de chaque bactérie elle-même. Quand elles sont injectées ensemble dans un animal, la souris meurt d'une infection par *Pneumococcus*. Des bactéries L virulentes peuvent être récupérées post mortem dans la souris.

Dans les premières expériences, les bactéries L mortes étaient de type I. Les bactéries R vivantes dérivait du type II. Les bactéries virulentes récupérées de l'infection mixte avaient le manteau lisse du type I. Ainsi un facteur des bactéries L mortes de type I peut **transformer** les bactéries R vivantes pour qu'elles synthétisent le polysaccharide capsulaire de type I et deviennent ainsi virulentes.

Le composant des bactéries mortes responsable de la transformation fut appelé le **principe transformant**. La purification de ce composant nécessitait le développement d'un système acellulaire, dans lequel des extraits des bactéries L mortes pouvaient être ajoutés aux bacté-

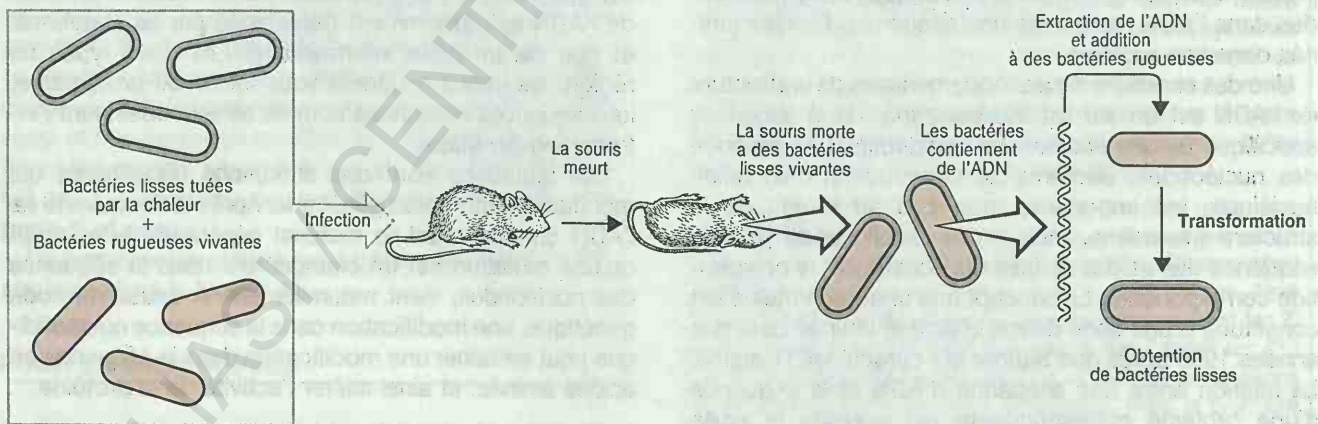


Figure 2.1

Le principe transformant est l'ADN.

Ni les bactéries lisses thermiquement détruites, ni les bactéries rugueuses (mutées) vivantes ne peuvent tuer les souris. Mais le mélange tue les souris ; et des bactéries lisses vivantes peuvent être retrouvées parmi celles-ci. La transformation de bactéries rugueuses inactives en bactéries lisses virulentes peut être réalisée *in vitro* par l'addition de l'ADN extrait des bactéries lisses.

ries R vivantes avant injection à l'animal. En 1944 Avery et ses collègues démontrèrent chimiquement que le principe transformant isolé était l'**acide désoxyribonucléique (ADN)**.

Le rôle de l'ADN fut confirmé avec la purification de la désoxyribonucléase (DNAase). Cette enzyme détruit spécifiquement l'ADN. L'addition de DNAase à la préparation de principe transformant empêche la transformation. La possibilité que la transformation pouvait être due à la présence d'une quelconque protéine contaminante présente dans la préparation d'ADN fut exclue en montrant que le traitement avec la trypsine (enzyme qui dégrade la plupart des protéines) n'avait aucun effet sur l'activité.

Le fait surprenant de ce résultat est qu'à cette époque l'ADN n'était même pas connu comme étant un composant de *Pneumococcus*, bien que l'on savait depuis de nombreuses décennies qu'il était un constituant majeur des chromosomes eucaryotes. En montrant que le matériel génétique d'un procaryote est l'ADN, ce résultat apporte alors une vue unifiée de la base de l'hérédité chez les bactéries et les organismes supérieurs.

La portée de ce résultat était précisément énoncée dans le papier original : « La substance inductrice, étant données ses propriétés chimiques et physiques, semble être une forme d'ADN hautement polymérisée et visqueuse. D'autre part, le polysaccharide capsulaire de type I, dont la synthèse est induite par cet agent transformant, est constitué principalement d'un polysaccharide non azoté... Il est donc évident que la substance inductrice et la substance produite sont chimiquement distinctes et ont des actions biologiques spécifiques et que toutes deux sont requises pour déterminer la spécificité des cellules auxquelles elles appartiennent ».

Cette discussion permet d'introduire une *distinction entre le matériel génétique et les produits de son expression*, distinction qui sera implicite dans les études suivantes.

L'ADN EST LE MATÉRIEL GÉNÉTIQUE (PRESQUE) UNIVERSEL

Après avoir montré que le principe transformant était de l'ADN, il fallait montrer dans un second temps que l'ADN constituait le matériel génétique dans un système assez différent. Le phage T2 est un virus qui infecte la bactérie *E. coli*. Quand les particules virales sont ajoutées aux bactéries, elles s'adsorbent sur la surface externe, du matériel entre dans la bactérie, et environ 20 minutes plus

tard chaque bactérie éclate (se lyse) et libère, en grand nombre, de nouvelles particules virales.

En 1952, Hershey et Chase infectèrent des bactéries avec des phages T2 marqués avec des isotopes radioactifs *soit* dans leur partie ADN (avec P^{32}), *soit* dans leur partie protéique (avec S^{35}). La **Figure 2.2** illustre les résultats de cette expérience. Les bactéries infectées ont été agitées dans un mélangeur et deux fractions ont été séparées par centrifugation. L'une contenait les enveloppes vides des phages, libérées de la surface des bactéries ; celles-ci sont constituées de protéines et contiennent donc le marquage radioactif S^{35} . L'autre fraction contenait les bactéries infectées elles-mêmes.

Le marquage P^{32} était présent majoritairement dans les bactéries infectées. Les nouvelles particules virales produites par l'infection contenaient environ 30 % du marquage P^{32} d'origine. La descendance a reçu très peu — moins de 1 % — des protéines contenues dans la population de départ. Cette expérience montrait donc directement que l'ADN des phages parents entre dans les bactéries et s'intègre en partie dans les nouveaux phages ; c'était exactement le mode de transmission attendu pour le matériel génétique.

La classification des virus comme « êtres vivants » a longtemps été matière à discussion. L'exemple classique du phage T2 montre que le virus est capable de réquisitionner la machinerie d'une cellule hôte infectée pour fabriquer d'autres copies de lui-même. Le phage possède un matériel génétique dont le comportement est analogue à celui des génomes cellulaires : ses caractéristiques sont fidèlement reproduites et elles sont soumises aux mêmes règles qui gouvernent l'hérédité. Le cas de T2 renforce donc la conclusion générale que le matériel génétique est l'ADN.

Bien qu'elles montrent clairement le rôle de l'ADN en tant que matériel génétique, les expériences d'infection avec les phages ne sont pas aussi précises que la transformation bactérienne pour exclure les protéines contaminant l'ADN. La purification poussée de l'ADN transformant réduit à des niveaux ridiculement bas la quantité de contaminant éventuellement présent. Pendant l'infection avec les phages, cependant, une petite quantité de protéines est en fait transférée dans les bactéries en même temps que l'ADN ; mais, évidemment, on ne retrouve que l'ADN dans la descendance.

Il est clair que le matériel génétique des bactéries et des phages est l'ADN. Mais qu'en est-il des eucaryotes ? Pendant longtemps l'évidence fut uniquement déductive. L'ADN est présent au bon endroit et se comporte de

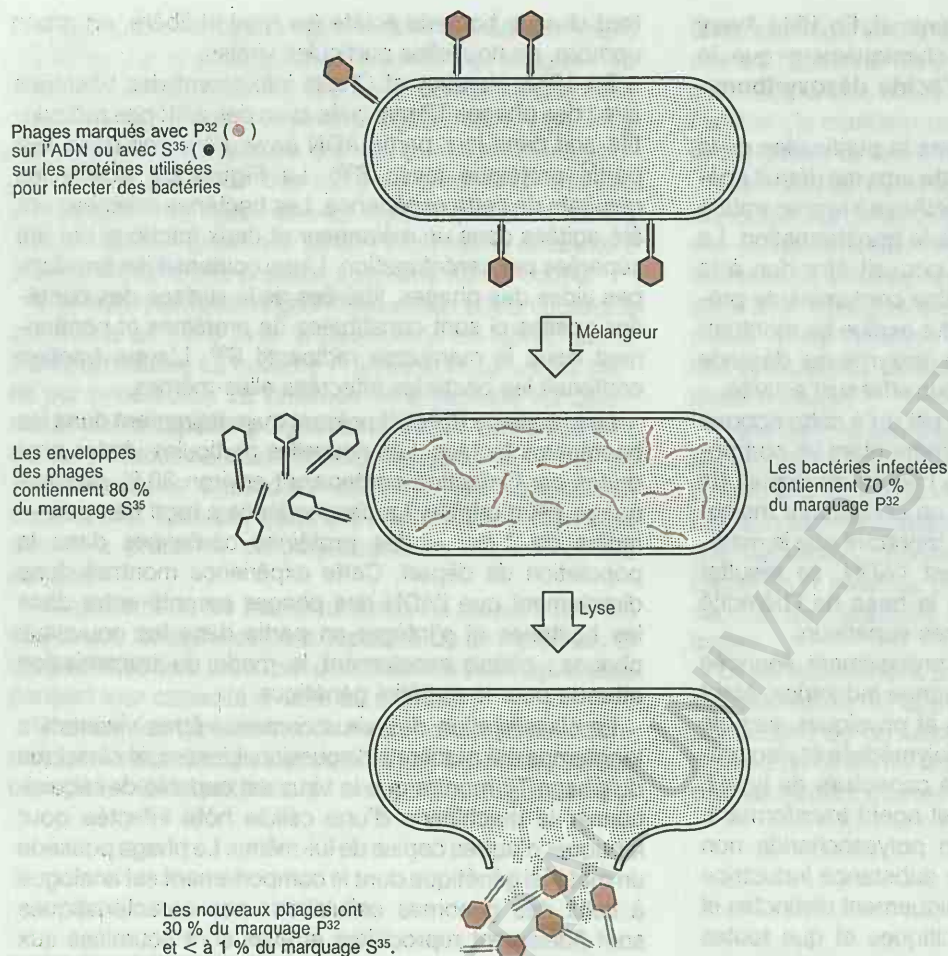


Figure 2.2

Le matériel génétique du phage T2 est l'ADN.

Quand des phages sont marqués au P^{32} sur l'ADN ou au S^{35} sur les protéines, le marquage P^{32} entre dans les bactéries infectées et peut être récupéré dans les nouveaux phages. Mais le marquage S^{35} n'est pas transmis aux descendants.

façon appropriée. On en eut la preuve directe seulement longtemps après que le fait fut considéré comme acquis.

En discutant ses résultats sur la transformation, Avery fit un commentaire dont les implications furent plus larges qu'il aurait pu imaginer : « si nous avons raison », écrivait-il, « cela signifie que les acides nucléiques ne sont pas uniquement des substances structurellement importantes mais sont des substances fonctionnellement actives pour déterminer les activités biochimiques et les caractéristiques spécifiques des cellules et que grâce à une subs-

tance chimique connue il est possible d'induire des changements prévisibles et héréditaires dans les cellules ». Il pensait au système bactérien, mais des résultats similaires ont maintenant été obtenus avec les eucaryotes.

Quand de l'ADN est ajouté à des populations de cellules eucaryotes isolées en culture, l'acide nucléique entre dans les cellules, et dans certaines d'entre elles permet la production de nouvelles protéines. D'abord réalisées avec de l'ADN extrait *en masse*, ces expériences peuvent maintenant être réalisées avec de l'ADN purifié

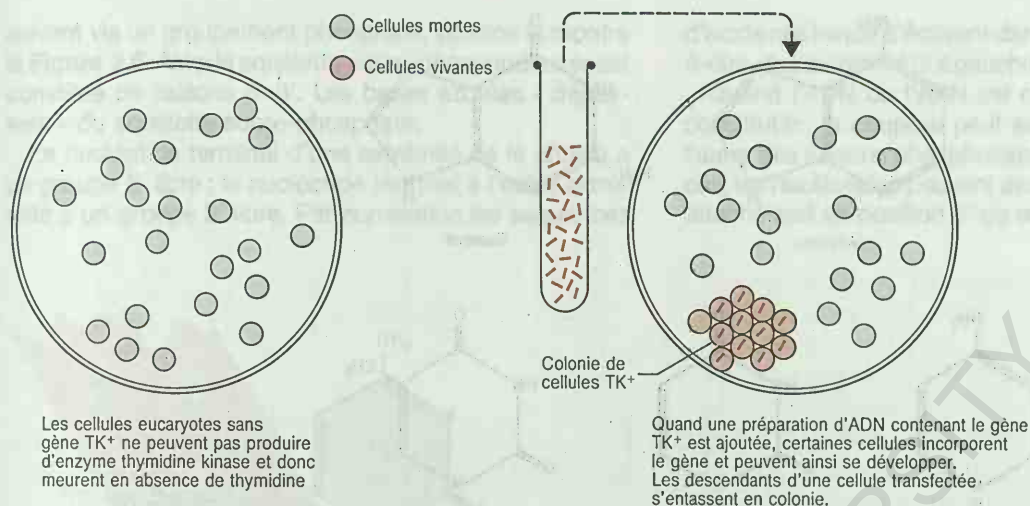


Figure 2.3

Les cellules eucaryotes peuvent acquérir un nouveau phénotype par transfection avec un ADN ajouté.

dont l'incorporation conduit à la production d'une protéine déterminée. La **Figure 2.3** décrit un des systèmes standard.

Bien que, pour des raisons historiques, ces expériences soient décrites comme une **transfection** quand elles sont réalisées avec des cellules eucaryotes, elles sont une contrepartie directe de la transformation bactérienne. L'ADN introduit dans la cellule réceptrice peut s'intégrer dans son matériel génétique, et être transmis comme n'importe quelle portion de ce matériel. Au début, ces expériences réussirent essentiellement avec des cellules isolées adaptées à se développer dans un milieu de culture. Depuis, cependant, l'ADN a été introduit dans des œufs de souris par micro-injection ; et il peut devenir une partie stable du matériel génétique de la souris (voir Chapitre 31).

De telles expériences montrent non seulement que l'ADN est le matériel génétique chez les eucaryotes, mais également qu'il peut être transféré entre des espèces différentes et rester fonctionnel.

Le matériel génétique de tous les organismes connus et de nombreux virus est l'ADN. Cependant, il existe également certains virus chez lesquels le matériel génétique est représenté par un autre acide nucléique, l'**acide ribonucléique (ARN)**. Bien que sa formule chimique soit légèrement différente de celle de l'ADN, dans certaines circonstances l'ARN exerce le même rôle. D'une façon générale la nature du matériel génétique est donc toujours un acide nucléique ; en fait c'est l'ADN sauf dans certains virus qui utilisent l'ARN à la place.

LES COMPOSANTS DE L'ADN

Le fait que les bases sont présentes en quantités différentes dans les ADN de différentes espèces, introduit le concept que *la séquence de bases pourrait être la forme sous laquelle l'information génétique est véhiculée*. Dès les années 1950, le concept d'information génétique était courant : le double problème posé était d'élaborer la structure de l'acide nucléique et d'expliquer comment une séquence de bases dans l'ADN pouvait représenter une séquence d'acides aminés dans une protéine.

Un acide nucléique est constitué d'une séquence de nucléotides liés chimiquement les uns aux autres. Chaque nucléotide contient un noyau hétérocyclique constitué d'atomes de carbone et d'azote (la **base azotée**), un sucre à cinq carbones sous forme cyclique (un **pentose**) et un groupe **phosphate**.

Les bases azotées constituent deux groupes montrés dans la **Figure 2.4** : les **pyrimidines** et les **purines**. Les pyrimidines ont un noyau à six côtés ; les purines ont des noyaux juxtaposés à cinq et six côtés. Chaque acide nucléique est synthétisé à partir de quatre types de bases uniquement. Les deux mêmes purines, adénine et guanine, sont présentes dans l'ADN et l'ARN. Les deux pyrimidines de l'ADN sont la cytosine et la thymine ; dans l'ARN l'uracile remplace la thymine. La seule différence entre l'uracile et la thymine est la présence d'un groupement méthyle en position C₅. Les bases sont habituellement désignées par leurs initiales ; ainsi l'ADN contient A, G, C, T alors que l'ARN contient A, G, C, U.

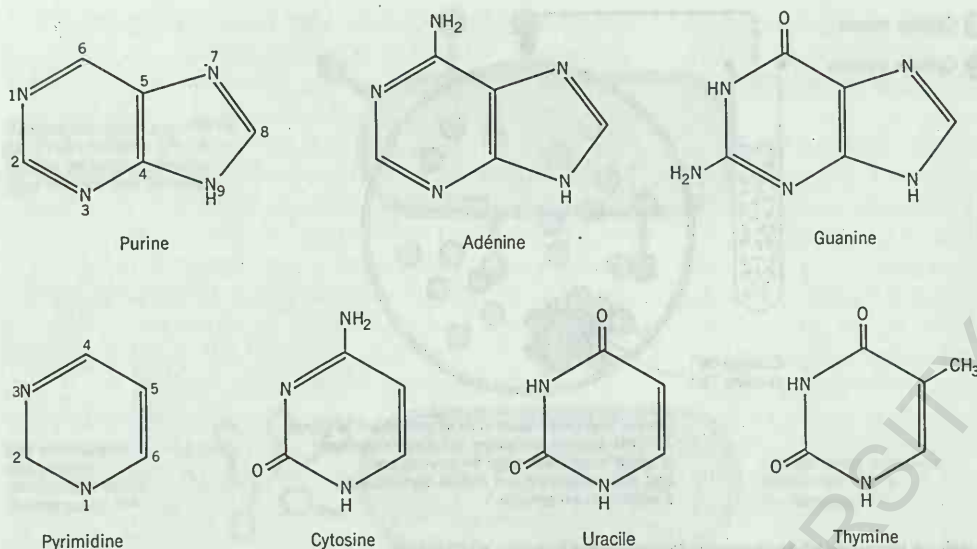


Figure 2.4

Purines et pyrimidines fournissent les bases azotées dans les acides nucléiques.

Les noyaux « purine » et « pyrimidine » montrent les structures générales de chaque type de base ; les chiffres identifient les positions sur le noyau.

Deux types de pentose sont présents dans les acides nucléiques. Ils différencient l'ADN et l'ARN et sont à l'origine des noms généraux des deux types d'acide nucléique. La Figure 2.5 montre leur structure. Dans l'ADN le pentose est le **2-désoxyribose** ; alors que dans l'ARN c'est un **ribose**. La différence réside dans l'absence/présence du groupement hydroxyle en position 2 sur le sucre.

La base azotée est reliée à la position 1 du pentose par une liaison glycosidique à partir du N¹ des pyrimidines ou du N⁹ des purines. Pour éviter toute ambiguïté entre les systèmes de numérotation des noyaux hétérocycliques et du sucre, les positions sur le pentose reçoivent un prime (').

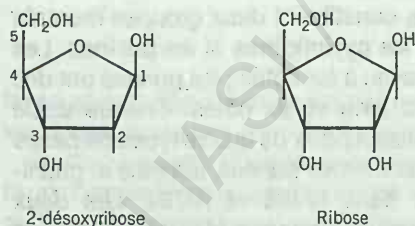


Figure 2.5

Le 2-désoxyribose est le sucre de l'ADN et le ribose est le sucre de l'ARN.

Les atomes de carbone sont numérotés comme indiqué pour le 2-désoxyribose.

Une base liée à un sucre est appelée un **nucléoside** ; quand un groupe phosphate est présent, l'ensemble base-sucre-phosphate est appelé **nucléotide**. La nomenclature des unités individuelles est décrite dans le Tableau 2.1.

Les nucléotides fournissent les éléments de construction à partir desquels les acides nucléiques sont formés. Les nucléotides sont liés ensemble dans une **chaîne polynucléotidique** pour former une ossature constituée d'une succession de sucre et de phosphate. La position 5' d'un pentose est connectée à la position 3' du pentose

Tableau 2.1

Les bases, les nucléosides et les nucléotides ont des noms apparentés.

Base	Nucléoside	Nucléotide	Abréviation	
			ARN	ADN
Adénine	adénosine	acide adénylique	AMP	dAMP
Guanine	guanosine	acide guanylique	GMP	dGMP
Cytosine	cytidine	acide cytidylique	CMP	dCMP
Thymine	thymidine	acide thymidylique		dTMP
Uracile	uridine	acide uridylique	UMP	

Les abréviations de la forme NMP se rapportent au nucléoside mono-phosphate ; « d » est utilisé pour indiquer la forme désoxy et son absence implique la présence d'un groupement 2'-OH.

suivant via un groupement phosphate, comme le montre la **Figure 2.6**. Ainsi le squelette sucre-phosphodiester est constitué de liaisons 5'-3'. Les bases azotées « dépassent » du squelette sucre-phosphate.

Le nucléotide terminal d'une extrémité de la chaîne a un groupe 5' libre ; le nucléotide terminal à l'autre extrémité a un groupe 3' libre. Par convention les séquences

d'acide nucléique s'écrivent dans la direction 5'-3' - c'est-à-dire, de l'extrémité 5' à gauche à l'extrémité 3' à droite.

Quand l'ADN ou l'ARN est cassé en ses nucléotides constitutifs, la coupure peut se réaliser sur un côté ou l'autre des liaisons phosphodiester. Selon les circonstances, les nucléotides peuvent avoir leur groupe phosphate attaché soit en position 5' ou en position 3' du pentose,

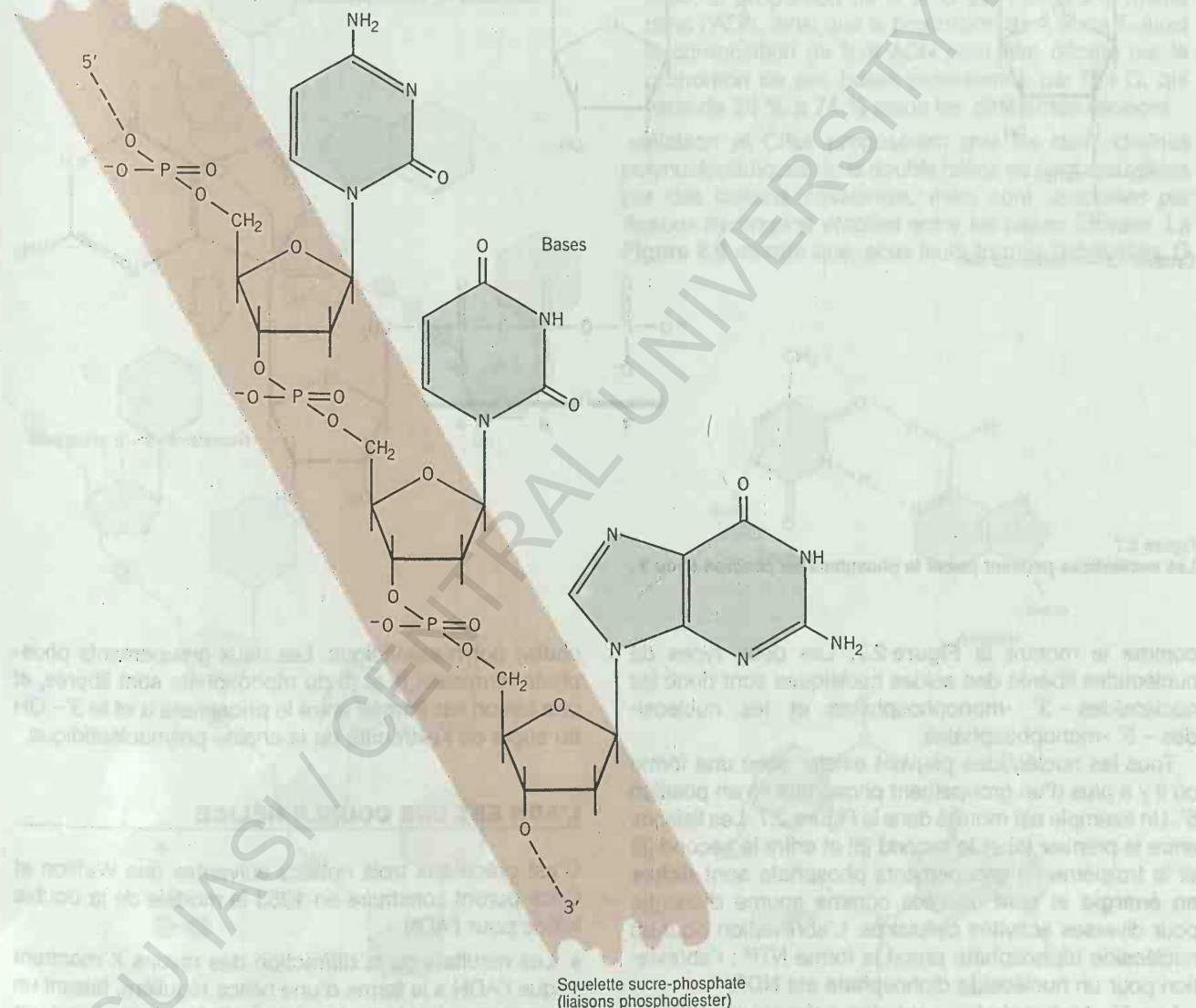


Figure 2.6

Une chaîne polynucléotidique est constituée d'une série de liaisons 5'-3' sucre-phosphate qui forment un squelette sur lequel les bases font saillie.

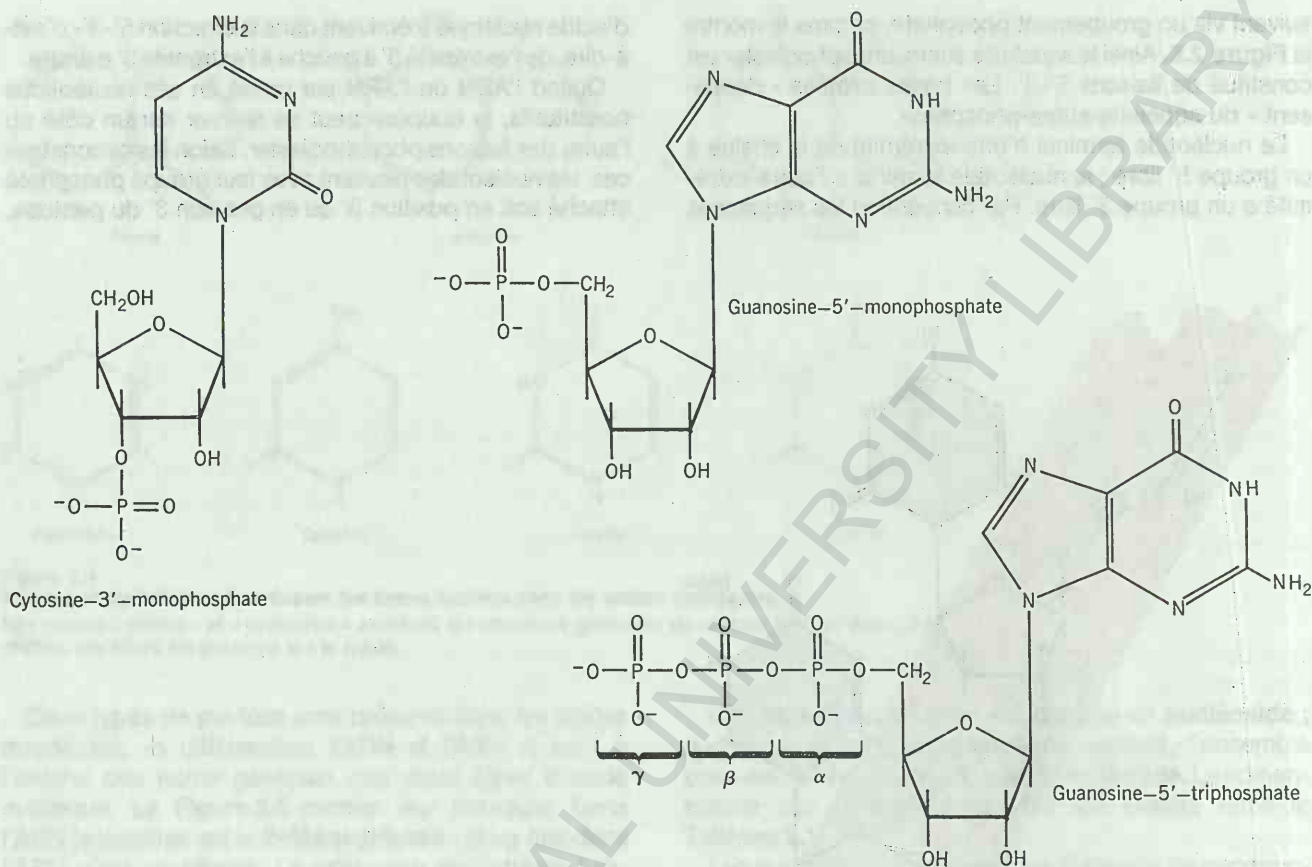


Figure 2.7

Les nucléotides peuvent porter le phosphate en position 5' ou 3'.

comme le montre la **Figure 2.7**. Les deux types de nucléotides libérés des acides nucléiques sont donc les nucléosides - 3' -monophosphates et les nucléosides - 5' -monophosphates.

Tous les nucléotides peuvent exister sous une forme où il y a plus d'un groupement phosphate lié en position 5'. Un exemple est montré dans la Figure 2.7. Les liaisons entre le premier (α) et le second (β) et entre le second (β) et le troisième (γ) groupements phosphate sont **riches en énergie** et sont utilisées comme source d'énergie pour diverses activités cellulaires. L'abréviation pour un nucléoside triphosphate prend la forme NTP ; l'abréviation pour un nucléoside diphosphate est NDP.

Les 5' -triphosphates sont des précurseurs pour la synthèse des acides nucléiques. La **Figure 2.8** montre la réaction dans laquelle l'extrémité 5' du triphosphate réagit avec un groupement 3'-OH de l'extrémité de la

chaîne polynucléotidique. Les deux groupements phosphate terminaux (γ et β) du triphosphate sont libérés, et une liaison est formée entre le phosphate α et le 3'-OH du sucre de l'extrémité de la chaîne polynucléotidique.

L'ADN EST UNE DOUBLE HÉLICE

C'est grâce aux trois notions suivantes que Watson et Crick purent construire en 1953 le modèle de la double hélice pour l'ADN.

- Les résultats de la diffraction des rayons X montrent que l'ADN a la forme d'une hélice régulière, faisant un tour complet tous les 34 Å (3,4 nm) et ayant un diamètre d'environ 20 Å (2 nm). Puisque la distance entre des nucléotides adjacents est de 3,4 Å, il doit donc y avoir 10 nucléotides par tour.

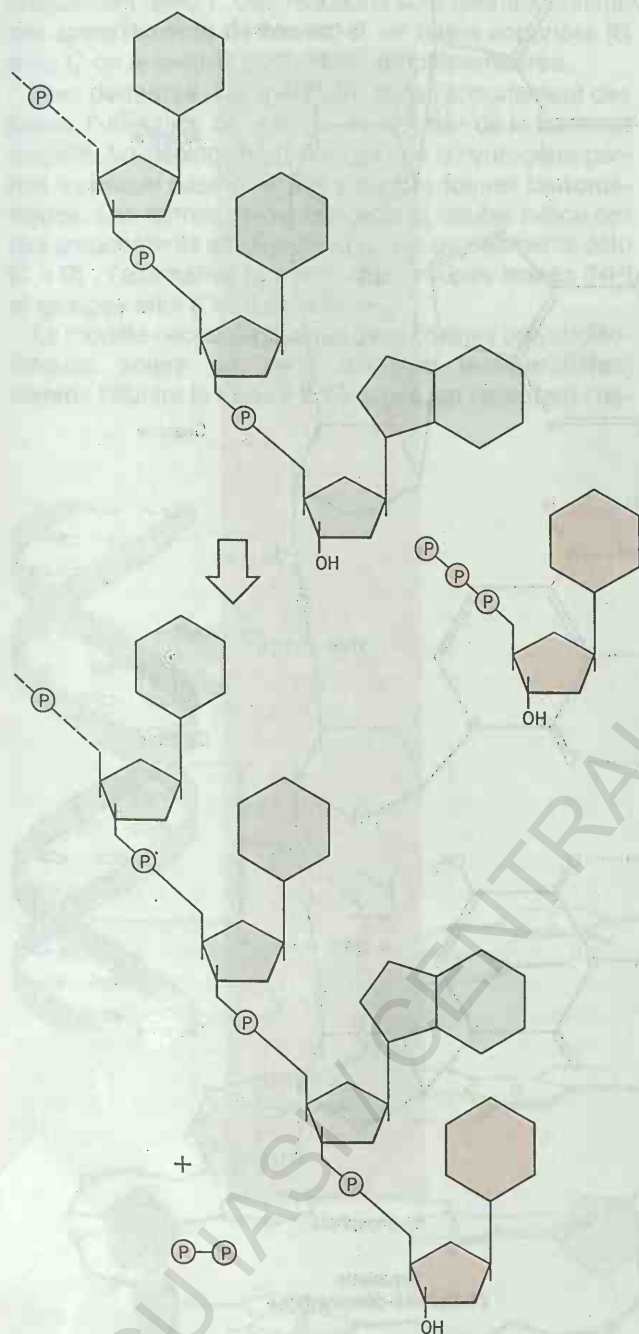


Figure 2.8
La synthèse des acides nucléiques résulte de l'addition de la partie nucléoside-5'-monophosphate d'un nucléoside triphosphate à l'extrémité 3'-OH de la chaîne polynucléotidique.

- La densité de l'ADN suggère que l'hélice doit contenir deux chaînes polynucléotidiques. Le diamètre constant de l'hélice peut s'expliquer si les bases de chaque chaîne sont tournées vers l'intérieur et sont limitées de façon à ce qu'une purine soit toujours opposée à une pyrimidine, évitant l'appariement purine-purine (structure trop volumineuse) ou pyrimidine-pyrimidine (structure trop petite).
- Indépendamment des quantités exactes de chaque base, la proportion de G et C est toujours la même dans l'ADN, ainsi que la proportion de A et de T. Ainsi la composition de tout ADN peut être décrite par la proportion de ses bases représentée par G + C, qui varie de 26 % à 74 % selon les différentes espèces.

Watson et Crick proposèrent que les deux chaînes polynucléotidiques de la double hélice ne sont pas reliées par des liaisons covalentes, mais sont *associées par liaisons hydrogène établies entre les bases azotées*. La **Figure 2.9** montre que, sous leurs formes habituelles, G

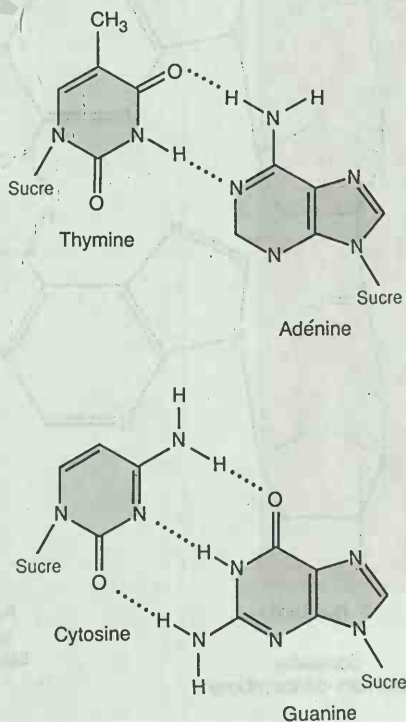


Figure 2.9
L'appariement complémentaire des bases implique la formation de deux liaisons hydrogène entre A et T, et de trois liaisons hydrogène entre G et C. Aucune autre paire ne se forme dans l'ADN.

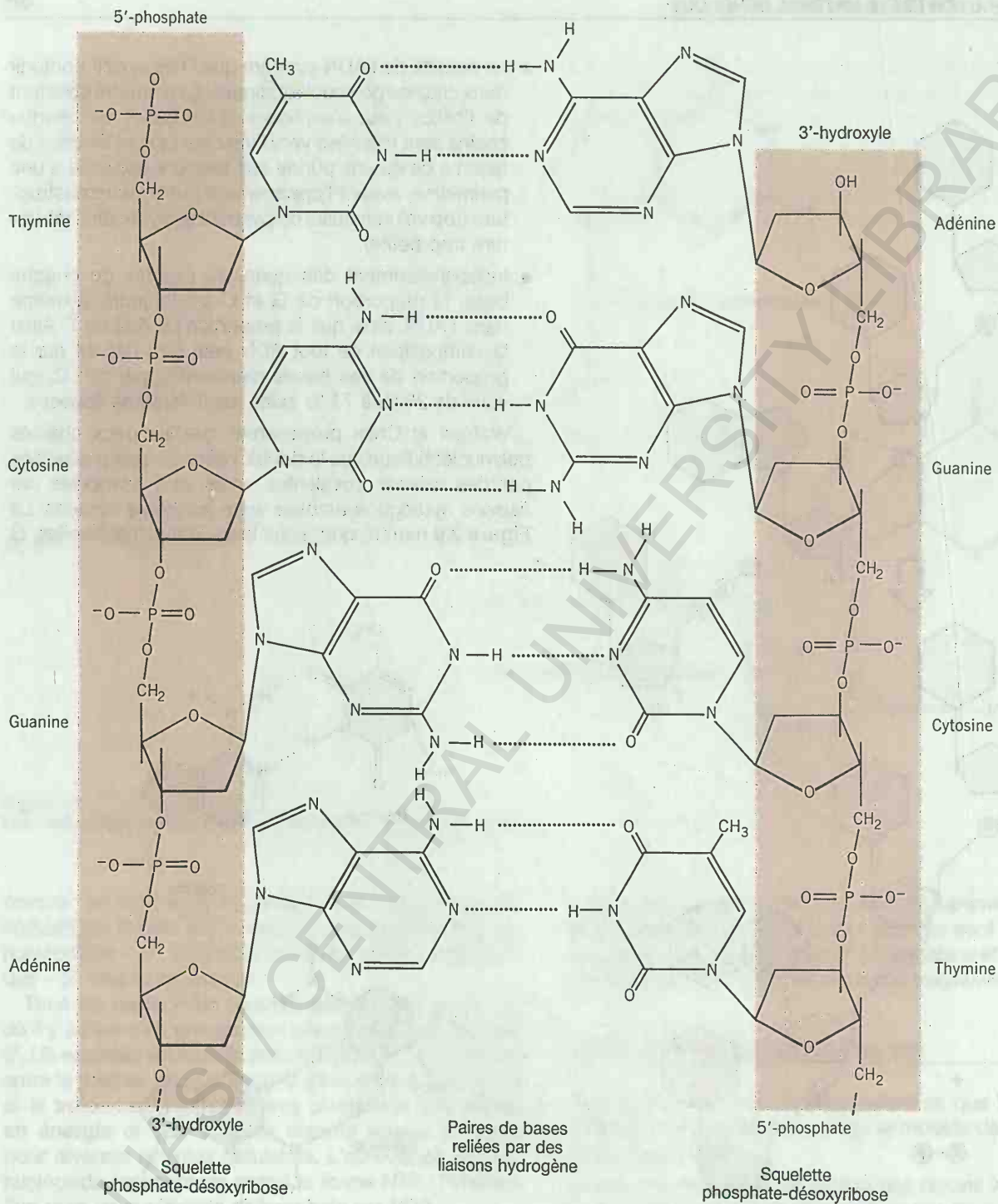


Figure 2.10

La double hélice a une largeur constante parce que les purines font toujours face aux pyrimidines dans les paires de bases complémentaires A-T et G-C.

En lisant la page de haut en bas, le brin gauche va de 5' en 3' et le brin droit va de 3' en 5'. La figure est schématique et ne montre pas l'enroulement des deux brins l'un par rapport à l'autre.

peut établir une liaison hydrogène spécifique seulement avec C, alors que A peut se lier de façon spécifique uniquement avec T. Ces réactions sont décrites comme des **appariements de bases**, et les bases appariées (G avec C ou A avec T) sont dites **complémentaires**.

Pour permettre une spécificité dans l'appariement des bases, l'utilisation de la forme appropriée de la base est cruciale. Le déplacement d'un atome d'hydrogène permet à chaque base d'exister sous des formes **tautomériques**. Les formes présentes dans la double hélice ont des groupements aminés (NH_2) et des groupements céto ($\text{C}=\text{O}$) ; l'alternative tautomérique groupes iminés (NH) et groupes enol (COH) est exclue.

Le modèle nécessite que les deux chaînes polynucléotidiques soient en sens opposés (**antiparallèles**) comme l'illustre la **Figure 2.10**. Donc, en regardant l'hé-

lice, une chaîne va dans le sens 3'-5' alors que sa partenaire va dans le sens 5'-3'.

Le squelette sucre-phosphate est à l'extérieur et porte des charges négatives sur les groupements phosphate. Quand l'ADN est en solution *in vitro*, les charges sont neutralisées par liaison avec des ions métalliques ; généralement avec Na^+ . A l'état naturel *in vivo*, les protéines chargées positivement fournissent un élément de neutralisation. Ces protéines peuvent jouer un rôle important en déterminant l'organisation de l'ADN dans la cellule.

Les bases sont disposées à l'intérieur. Ce sont des structures planes, disposées par paires perpendiculairement à l'axe de l'hélice. Si on imagine la double hélice comme un escalier en spirale, les paires de bases en forment les marches, comme l'illustre schématiquement la **Figure 2.11**. En progressant le long de l'hélice, les

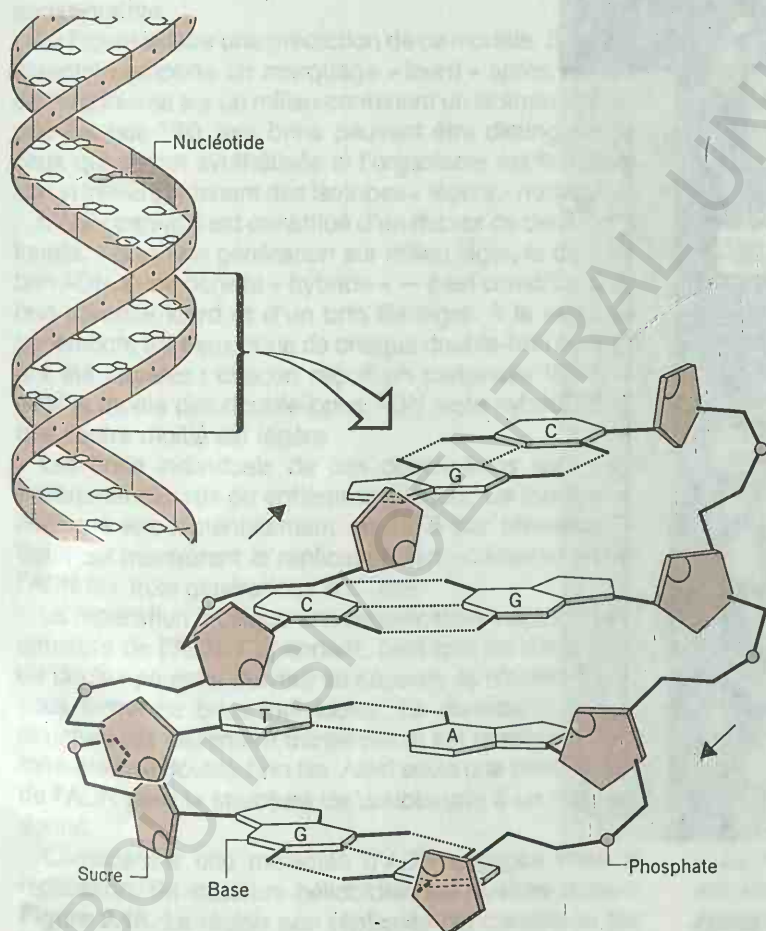


Figure 2.11

Les paires de bases sont perpendiculaires au squelette sucre-phosphate.

bases sont empilées les unes au-dessus des autres, comme une pile d'assiettes.

Les paires de bases contribuent à la stabilité thermodynamique de la double hélice de deux façons. L'énergie est libérée :

- par les liaisons hydrogène qui s'établissent entre les bases à l'intérieur de chaque paire ;
- et par l'**entassement hydrophobe des bases**, résultant des interactions entre les systèmes électroniques des paires de bases entassées.

Chaque paire de bases pivote d'environ 36° autour de l'axe de l'hélice par rapport à la paire de bases suivante. Ainsi environ 10 paires de bases font un tour complet de 360° . La torsion des deux brins l'un autour de l'autre forme une double hélice avec un **petit sillon** (environ $12 \text{ \AA}$) et un **grand sillon** (environ $22 \text{ \AA}$) comme on peut le voir dans le modèle présenté à l'échelle dans la **Figure 2.12**. La double hélice est **droite** ; elle tourne dans le sens des aiguilles d'une montre par rapport à l'axe. Ces caractéristiques définissent la **forme B** de l'ADN.

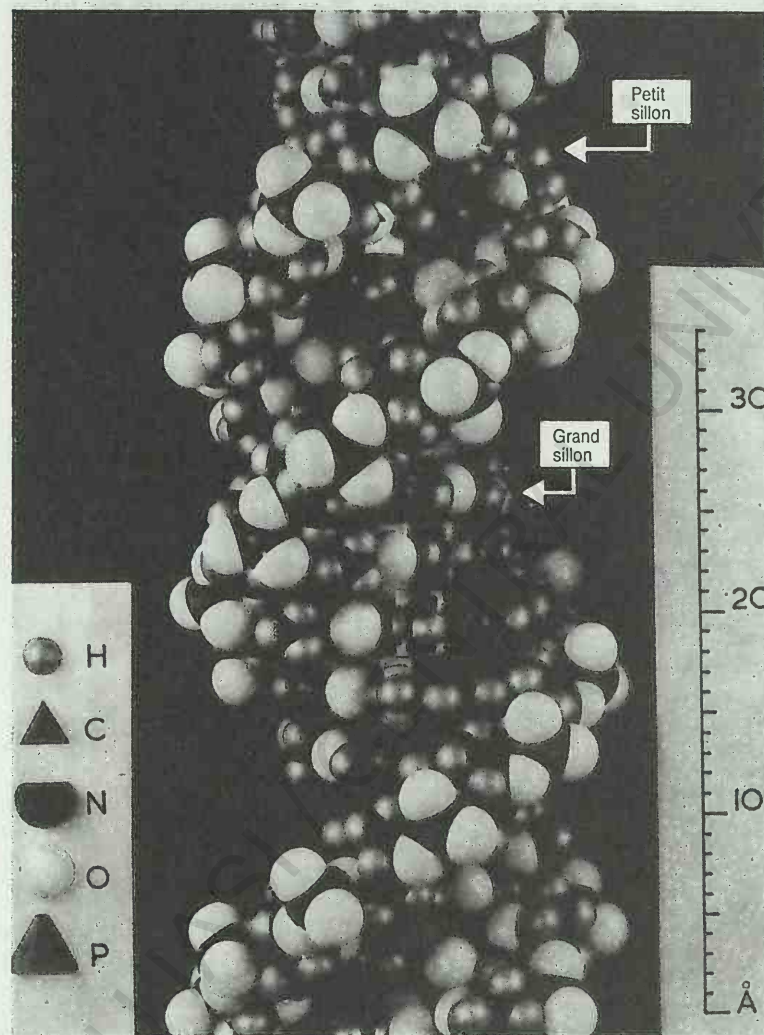


Figure 2.12

Les deux brins de l'ADN forment une double hélice.

Photographie du modèle spatial aimablement fournie par Maurice Wilkins.

LA RÉPLICATION DE L'ADN EST SEMI-CONSERVATIVE

Il est crucial que le matériel génétique soit reproduit avec exactitude. Puisque les deux brins polynucléotidiques sont réunis uniquement par des liaisons hydrogène, ils peuvent se séparer. La spécificité de l'appariement des bases suggère que chaque brin pourrait se comporter comme un **modèle** pour la synthèse d'un brin complémentaire, conformément à la description de la **Figure 2.13**. *La structure de l'ADN comporte donc l'information nécessaire pour perpétuer sa séquence.*

La **Figure 2.14** illustre les résultats de ce mode de réplication. Le **double-brin parental** est répliqué pour former deux **double-brins fils**, constitués chacun d'un brin parental et d'un brin fils (nouvellement synthétisé). Ainsi l'unité qui se conserve d'une génération à l'autre est représentée par un des deux brins de double-brin parental. Ce fonctionnement est appelé **réplication semi-conservative**.

La Figure illustre une prédiction de ce modèle. Si l'ADN parental comporte un marquage « lourd » après culture de l'organisme sur un milieu contenant un isotope approprié (tel que ^{15}N), ses brins peuvent être distingués de ceux qui seront synthétisés si l'organisme est transféré sur un milieu contenant des isotopes « légers » normaux.

L'ADN parental est constitué d'un duplex de deux brins lourds. Après une génération sur milieu léger, le double-brin ADN a une densité « hybride » — il est constitué d'un brin parental lourd et d'un brin fils léger. A la seconde génération, les deux brins de chaque double-brin hybride ont été séparés ; chacun reçoit un partenaire léger, et ainsi la moitié des double-brins ADN reste hybride alors que l'autre moitié est légère.

Les brins individuels de ces double-brins sont tous entièrement lourds ou entièrement légers. Ce modèle fut confirmé expérimentalement en 1958 par Meselson et Stahl qui montrèrent la réplication semi-conservative de l'ADN sur trois générations d'*E. coli*.

La réplication implique une modification majeure de la structure de l'ADN. Cependant, bien que les deux brins du duplex parental doivent se séparer, ils n'existent pas sous forme de brins individuels. La dissociation de la structure est seulement transitoire et est reversée dès la formation du double-brin fils. Ainsi seule une petite partie de l'ADN perd la structure de double-brin à un moment donné.

Considérons une molécule d'ADN engagée dans la réplication. Sa structure hélicoïdale est illustrée dans la **Figure 2.15**. La région non répliquée est constituée par le double-brin parental et s'ouvre sur la région répliquée

où les deux double-brins fils ont été formés. La structure en double hélice est rompue à la jonction des deux régions, que l'on appelle la **fourche de réplication**. La réplication implique la progression de la fourche de réplication le long de l'ADN parental, et ainsi il y a déroulement continu des brins parentaux et ré-enroulement des double-brins fils.

LE CODE GÉNÉTIQUE EST LU SOUS FORME DE TRIPLETS

Une séquence nucléotidique doit contenir assez d'unités codantes pour représenter les 20 acides aminés. Puisqu'il n'y a que 4 bases dans l'ADN, il faut donc utiliser plus d'une base pour déterminer chaque acide aminé.

Si 2 bases codaient pour chaque acide aminé, l'ADN pourrait déterminer $4^2 = 16$ acides aminés différents. Pour déterminer 20 acides aminés le rapport codant doit donc être supérieur ou égal à 3. Si le code génétique est un triplet, trois bases adjacentes doivent représenter chaque acide aminé. Puisqu'il y a $4^3 = 64$ combinaisons possibles, certains des triplets doivent être incapables de coder pour un acide aminé ou certains acides aminés doivent être codés par plus d'un triplet.

Le code est déchiffré par un appareillage complexe situé entre l'acide nucléique et la protéine (et qui est donc nécessaire si l'information portée par l'ADN doit être exprimée). Dans n'importe quelle région, un seul des deux brins de l'ADN code pour une protéine, c'est pourquoi nous écrivons le code génétique comme une séquence de bases (plutôt que de paires de bases).

Le code génétique est lu par groupes de trois nucléotides, chaque groupe représentant un acide aminé. Chaque séquence trinucéotidique est appelée **codon**. Un gène englobe un ensemble de codons qui est lu en série d'un point de départ délimitant une extrémité jusqu'à un point d'arrêt délimitant l'autre extrémité. Écrite dans le sens conventionnel $5' \rightarrow 3'$, la séquence nucléotidique du brin d'ADN codant pour une protéine correspond à la séquence des acides aminés de la protéine écrite dans le sens N-terminal vers C-terminal.

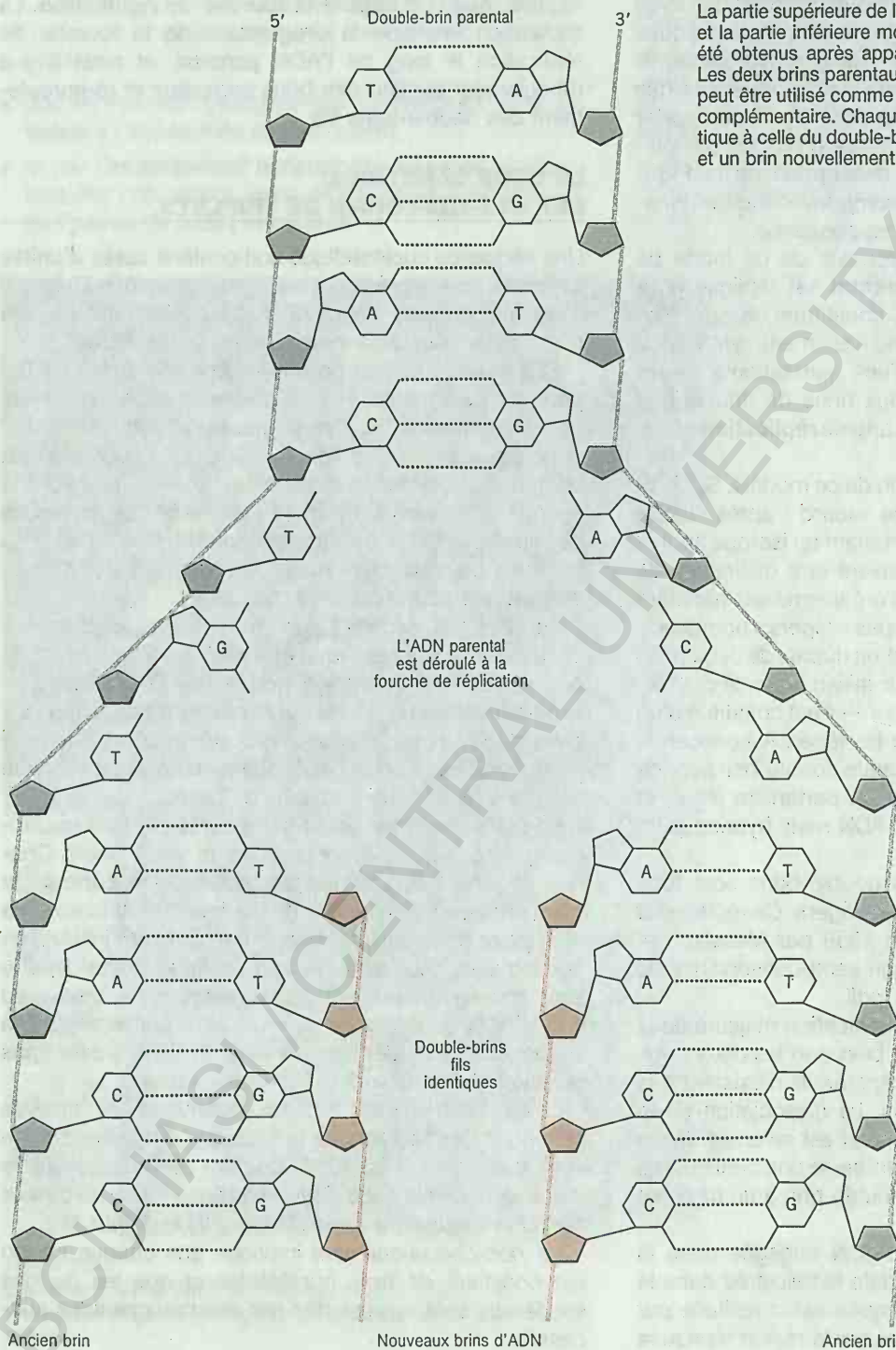
La base générale du code fut découverte par l'analyse génétique des mutants de la région *rII* d'un virus bactérien, le phage T4. En 1961, Crick et ses collaborateurs montrèrent que le code doit être lu sous forme de triplets *non chevauchants à partir d'un point de départ fixe*.

Le *non-chevauchement* implique que chaque codon est constitué de trois nucléotides et que les codons successifs sont représentés par des trinucéotides successifs.

Figure 2.13

L'appariement des bases fournit le mécanisme nécessaire à une réplication fidèle de l'ADN.

La partie supérieure de la figure montre un double-brin parental et la partie inférieure montre les deux double-brins fils qui ont été obtenus après appariement des bases complémentaires. Les deux brins parentaux ont été séparés de sorte que chacun peut être utilisé comme une matrice pour la synthèse d'un brin complémentaire. Chaque double-brin fils a une séquence identique à celle du double-brin parental et contient un brin parental et un brin nouvellement synthétisé.



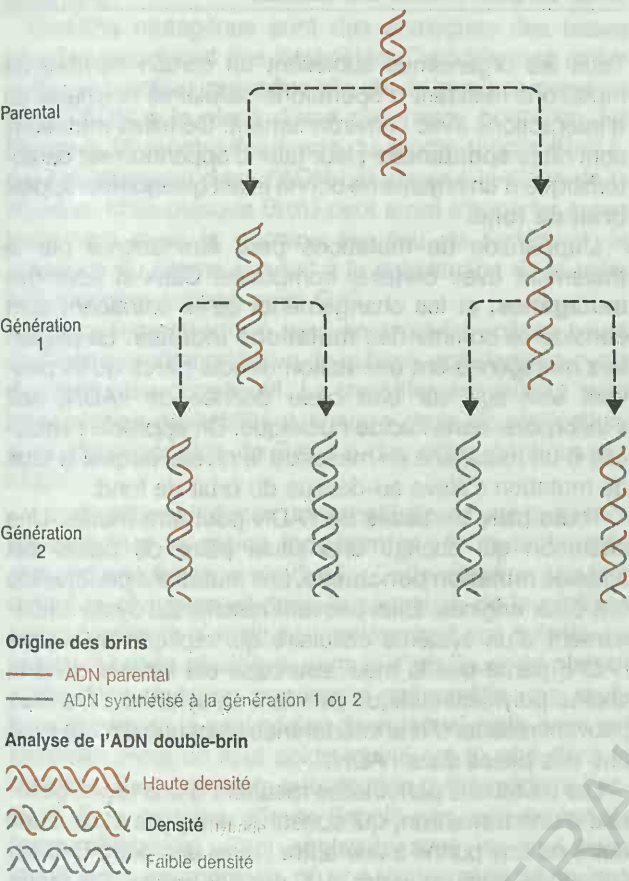


Figure 2.14 La réplication de l'ADN est semi-conservative.

L'utilisation d'un *point de départ fixe* signifie que l'assemblage d'une protéine doit commencer à une extrémité et évoluer jusqu'à l'autre, de sorte que les différentes parties de la séquence codante ne peuvent pas être lues indépendamment.

Si le code génétique est lu sous forme de triplets non chevauchants, il y a trois possibilités pour traduire une séquence nucléotidique en protéine, selon le point de départ choisi. Celles-ci sont appelées **cadres de lecture**. Pour la séquence

ACGACGACGACGACGACGACG

les trois cadres possibles de lecture sont

ACG ACG ACG ACG ACG ACG
CGA CGA CGA CGA CGA CGA
GAC GAC GAC GAC GAC GAC

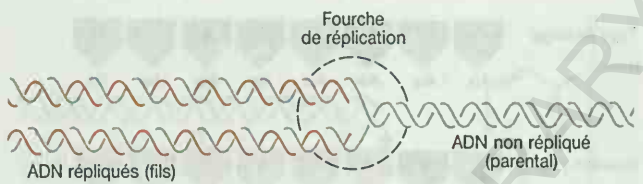


Figure 2.15 La fourche de réplication est la région de l'ADN où il y a transition entre l'ADN parental déroulé et les double-brins fils nouvellement répliqués.

Une mutation qui insère ou qui délète une seule base changera le cadre de lecture pour toute la séquence qui suit. Un changement de ce type est appelé **décalage du cadre** (frameshift). Puisque la séquence du nouveau cadre de lecture est complètement différente de l'ancien, la séquence en acides aminés de la protéine est totalement modifiée à partir du site de la mutation. Ainsi il est vraisemblable que la fonction de la protéine soit totalement perdue. Ce type de mutation est induit par les **acridines**, composés qui se fixent à l'ADN et qui tordent la structure de la double hélice, provoquant une incorporation de bases supplémentaires ou une omission pendant la réplication. Tout événement mutagène dû à une acridine conduit à l'addition ou au retrait d'une seule paire de bases.

Si un mutant acridine est produit par, disons, addition d'un nucléotide, il devrait retourner au type sauvage par délétion du nucléotide. Mais la réversion peut également être provoquée par la délétion d'une base différente, située en un site proche du premier. La seconde mutation est dite **suppressiv**e. (Dans le cadre de ce travail, « suppressive » est utilisé dans un sens inhabituel ; voir plus loin).

La Figure 2.16 montre que la combinaison d'une mutation acridine et de son suppresseur donne un cadre de lecture incorrect uniquement entre les deux mutations ; de chaque côté, la lecture est réalisée dans le bon cadre. Quand une recombinaison génétique est utilisée pour séparer les mutants acridine et leurs suppresseurs, le suppresseur peut être caractérisé comme une mutation différente.

Toutes les mutations acridine peuvent être classées dans l'un des deux ensembles, décrits comme (+) et (-). Chaque type de mutation entraîne un décalage du cadre, le type (+) par l'addition d'une base, le type (-) par la délétion d'une base. Les combinaisons de doubles mutations de type (+ +) et (- -) continuent à avoir un comportement de mutant. Mais les doubles mutations de types (+ -) ou (- +) se suppriment mutuellement à l'intérieur d'une même combinaison.

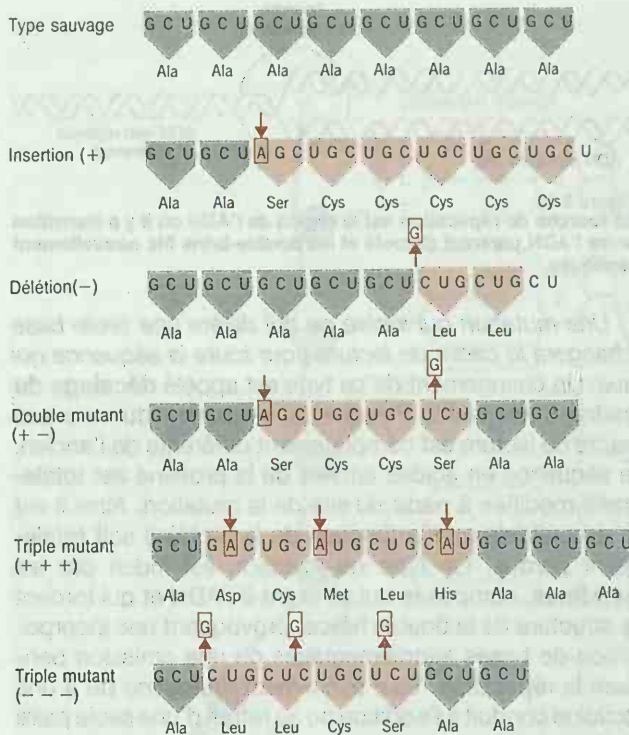


Figure 2.16

Les mutations avec décalage du cadre de lecture montrent que le code génétique est lu sous forme de triplets à partir d'un point de départ fixe.

Le cadre de lecture (triplet) est indiqué par des pentagones ; le gris indique les codons de type sauvage et la couleur indique les codons mutants. Les bases insérées ou délétées dans la séquence de type sauvage sont indiquées par des rectangles et des flèches.

Le code génétique doit être lu comme une séquence incluse dans un cadre de lecture fixé par le point de départ, de sorte que les additions ou les délétions se compensent, alors que les doubles additions ou les doubles délétions restent mutantes. Mais ceci ne révèle pas le nombre de nucléotides dont est constitué chaque codon.

Quand des triples mutants sont construits, seules les combinaisons (+ + +) et (- - -) ont le phénotype sauvage, alors que les autres combinaisons restent mutantes. Si nous considérons que trois additions ou trois délétions correspondent respectivement à l'addition ou à l'omission en tout, d'un seul acide aminé, ceci implique que le code est lu sous forme de triplets. Comme avec les doubles mutants compensateurs, une séquence d'acides aminés incorrecte est trouvée dans l'intervalle compris entre les sites de mutations alors que la séquence située de chaque côté reste de type sauvage, comme indiqué dans la Figure 2.16.

LES MUTATIONS PONCTUELLES CHANGENT UNE SEULE PAIRE DE BASES

Tous les organismes subissent un certain nombre de mutations résultant d'opérations cellulaires normales ou d'interactions avec l'environnement. De telles mutations sont dites **spontanées** ; leur taux d'apparition est caractéristique d'un organisme donné et est quelquefois appelé **bruit de fond**.

L'apparition de mutations peut être accrue par le traitement avec certains composés. Ceux-ci sont des **mutagènes**, et les changements qu'ils entraînent sont considérés comme des **mutations induites**. La plupart des mutagènes ont une action directe parce qu'ils peuvent soit agir sur une base donnée de l'ADN, soit s'incorporer dans l'acide nucléique. On apprécie l'efficacité d'un mutagène en mesurant le niveau auquel le taux de mutation s'élève au-dessus du bruit de fond.

Toute paire de bases de l'ADN peut être mutée. Une altération qui change une seule paire de bases est appelée **mutation ponctuelle**. Les mutations ponctuelles ont deux origines. Elles peuvent résulter du dysfonctionnement d'un système cellulaire qui réplique ou répare l'ADN, parce que la mauvaise base est insérée dans la chaîne polynucléotidique pendant sa synthèse. Ou elles peuvent résulter d'une interférence chimique directe avec une des bases dans l'ADN.

Les mutations ponctuelles résultent d'une façon générale d'une **transition**, qui substitue une pyrimidine à une autre ou une purine à une autre ; ainsi une paire G-C est échangée par une paire A-T ou vice versa. De façon moins courante elles peuvent résulter d'une **transversion**, où une purine est remplacée par une pyrimidine ou vice versa, de sorte qu'une paire A-T devient une paire T-A ou C-G.

La **conversion chimique** directe d'une base en une autre est une des sources de transition. L'acide nitreux provoque une désamination oxydative qui transforme la cytosine en uracile. La Figure 2.17 en montre les conséquences. Dans le cycle de réplication suivant, le U s'apparie avec le A et non avec le G avec lequel le C d'origine se serait apparié. Ainsi la paire CG est remplacée par une paire TA quand le A s'apparie avec le T au prochain cycle de réplication. L'acide nitreux désamine également l'adénine, entraînant la transition reverse de A-T en G-C.

Le **mauvais appariement des bases** est une autre cause de transition, quand des partenaires inhabituels s'apparient au mépris de la restriction usuelle sur l'appariement énoncée par Watson et Crick. Le mauvais appariement des bases se manifeste habituellement comme

une aberration résultant de l'introduction d'une base anormale.

Certains mutagènes sont des analogues des bases usuelles possédant des propriétés d'appariement ambiguës ; leur action mutagène résulte de leur incorporation dans l'ADN à la place d'une des bases normales. La **Figure 2.18** montre l'exemple du bromo-uracile (BrdU), qui est incorporé dans l'ADN par erreur à la place de la thymine. Mais puisque BrdU peut aussi s'apparier assez facilement avec la guanine (au lieu de l'adénine), la présence du brome conduit à la substitution de la paire A-T d'origine par une paire G-C.

L'appariement erroné peut se produire soit au cours de l'incorporation primitive de la base, soit dans un cycle de réplication consécutif. La transition est induite avec une certaine probabilité à chaque cycle de réplication, ainsi l'incorporation de BrdU a des effets continus sur l'ADN.

Les mutations induites par substitution d'une base sont souvent plus ou moins **inopérantes** : le mutant garde une certaine fonction résiduelle. Cette situation survient quand le changement dans la séquence de la protéine correspondante n'abolit pas entièrement son activité. La nature du code génétique permet d'expliquer ce phénomène. Une mutation ponctuelle qui modifie une seule base changera le seul codon dans lequel cette base est localisée. Ainsi un seul acide aminé est touché dans la protéine. Bien que cette substitution puisse réduire l'activité de la protéine, elle ne l'abolit pas entièrement de façon obligatoire. Ceci contraste avec les mutations induites par les acridines, où une longue série de codons peut être altérée par un décalage du cadre de lecture, et où la fonction protéique a complètement disparu.

On a longtemps pensé que les mutations ponctuelles constituaient le principal changement à l'intérieur des gènes. Cependant, on sait maintenant que des **insertions** de morceaux de matériel supplémentaire peuvent être assez fréquentes. La source du matériel inséré se trouve avec des **éléments transposables**, séquences d'ADN qui peuvent se déplacer d'un site à un autre. Une telle insertion détruit l'activité d'un gène. Là où ces insertions se sont manifestées, des **délétions** d'une partie ou de tout le matériel inséré, et quelquefois des régions adjacentes, peuvent se produire par la suite.

Il existe une différence significative entre les mutations ponctuelles et les insertions/délétions : la fréquence d'une mutation ponctuelle peut être accrue par les agents mutagènes, alors que l'apparition des changements dus à des éléments transposables est indifférente à ces composés. Cependant insertions et délétions peuvent également apparaître par d'autres mécanismes — par

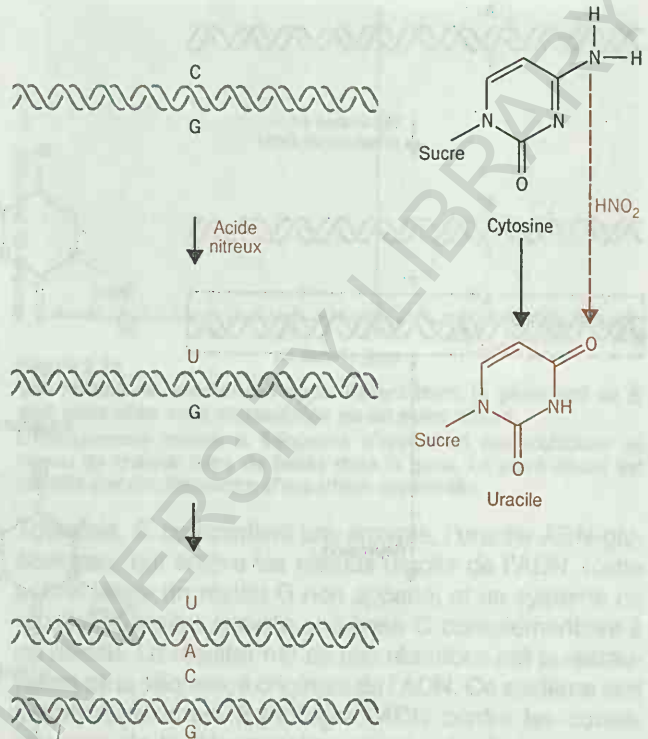


Figure 2.17

Les mutations peuvent être induites par la modification chimique d'une base.

L'acide nitreux désamine de façon oxydative la cytosine en uracile. La réplication engendre un double-brin fils avec la paire C-G de type sauvage ; mais l'autre possède une paire U-A, qui sera répliquée en paire T-A aux générations suivantes.

exemple, ceux impliquant des erreurs faites pendant la réplication ou la recombinaison — bien que ceux-ci soient probablement moins fréquents. Les changements induits par les acridines constituent aussi des insertions et des délétions.

LES MUTATIONS SONT CONCENTRÉES EN DES POINTS CHAUDS

Jusqu'à maintenant nous avons considéré les mutations comme des changements individuels dans la séquence de l'ADN qui influencent l'activité de l'unité génétique dans laquelle elles apparaissent. Si on considère les mutations en termes d'inactivation du gène, la plupart des gènes à l'intérieur d'une espèce ont des taux de mutation plus ou moins similaires en relation avec leur taille (c'est-à-dire, relativement à la cible pour la mutation). Mais si on considère les sites de mutation dans la séquence de l'ADN, on peut se demander si toutes les paires de bases dans un gène sont également suscepti-

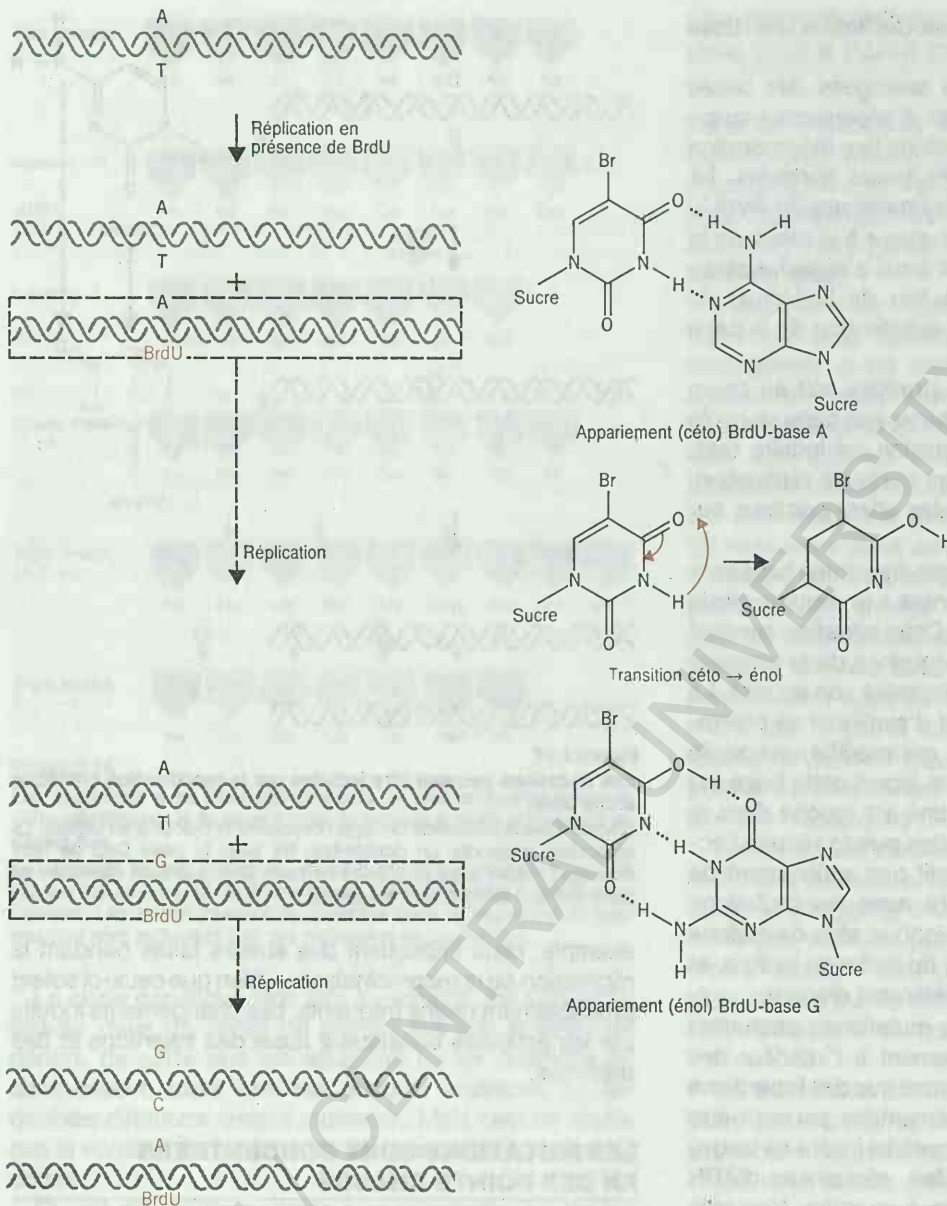


Figure 2.18

Les mutations peuvent être induites par l'incorporation de bases analogues dans l'ADN.

Le bromo-uracile contient un atome de brome à la place du groupement méthyle de la thymine. Sa présence permet à la transition céto - énol d'apparaître plus fréquemment, permettant à BrdU de s'apparier aussi bien avec la guanine qu'avec l'adénine. Ainsi si BrdU est incorporé dans l'ADN à la place de la thymine, il peut entraîner le remplacement de A par G avec une certaine fréquence à chaque génération. Il en résulte un remplacement des paires A-T par des paires G-C.

bles ou si certaines sont plus sujettes à être mutées que d'autres.

Cette question fut abordée en isolant un grand nombre de mutations indépendantes dans le même gène. De

nombreux mutants dont chacun a subi un événement mutationnel individuel sont obtenus. Ensuite le site de chaque mutation est identifié. La plupart des mutations se trouveront à des sites différents, mais certaines seront

localisées à la même position. Deux mutations isolées indépendamment au même site peuvent constituer exactement le même changement dans l'ADN (auquel cas le même événement est apparu à plus d'une occasion), ou elles peuvent constituer des changements différents (trois mutations ponctuelles différentes sont possibles à chaque paire de bases). La **Figure 2.19** montre la distribution des mutations dans le gène *lacI* de *E. coli*.

La probabilité statistique que plus d'une mutation apparaisse à un site particulier est donnée par la loi des tirages au hasard (comme on peut le voir dans la distribution de Poisson). Ainsi certains sites auront une, deux ou trois mutations alors que d'autres n'en auront aucune. Mais certains sites reçoivent beaucoup plus de mutations que le nombre attendu dans le cas d'une distribution au hasard ; ils peuvent avoir 10 ou 100 fois plus de mutations que prévu par des tirages aléatoires. Ces sites sont appelés **points chauds**. Les points chauds ne sont pas universels pour tous les types de mutations ; et des agents mutagènes différents peuvent avoir des points chauds différents.

Une cause majeure de mutation spontanée chez *E. coli* a été attribuée à la présence d'une base inhabituelle dans l'ADN. En plus des quatre bases qui sont insérées dans l'ADN au moment de sa synthèse, des **bases modifiées** sont parfois trouvées. Leur nom reflète leur origine ; elles sont produites par modification chimique d'une des quatre bases déjà présentes dans l'ADN. La base modifiée la plus commune est la 5-méthylcytosine, produite par action d'une méthylase (enzyme) qui ajoute un groupement méthyle à une petite proportion des résidus cytosine (au niveau de sites spécifiques dans l'ADN).

Les sites qui contiennent la 5-méthylcytosine constituent des points chauds pour les mutations ponctuelles spontanées. Dans chaque cas, la mutation prend la forme d'une transition G-C en A-T. Dans les souches d'*E. coli* qui sont incapables d'effectuer la réaction de méthylation, ces points chauds n'existent pas.

La raison de l'existence des points chauds est que la 5-méthylcytosine se désamine spontanément avec une fréquence appréciable ; le remplacement d'un groupement amino par un groupement céto convertit la 5-méthylcytosine en thymine (les structures des pyrimidines sont décrites dans la Figure 2.4). La conversion crée une association G-T dépariée, dont la séparation à la réplication suivante produit une paire G-C de type sauvage et une paire A-T mutée.

La **Figure 2.20** compare les effets de la désamination sur la 5-méthylcytosine (rare) et sur la cytosine plus commune.

La désamination de la cytosine produit de l'uracile,

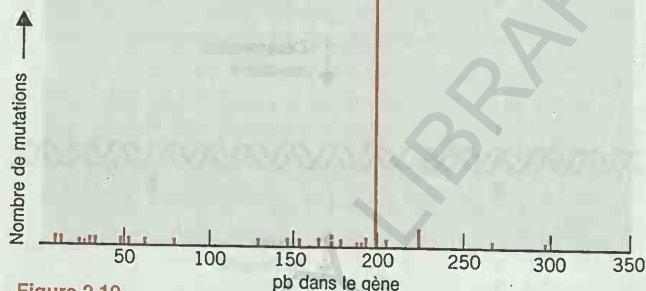


Figure 2.19

Des mutations spontanées apparaissent dans le gène *lacI* de *E. coli*, mais elles sont concentrées en un point chaud.

L'histogramme montre la fréquence d'apparition des mutations au niveau de chaque paire de bases dans le gène. Le point chaud est identifié par une fréquence d'apparition augmentée.

Toutefois, *E. coli* contient une enzyme, l'uracile-ADN-glycosidase, qui enlève les résidus uracile de l'ADN. Cette action laisse un résidu G non apparié, et un système de réparation insère ensuite une base C complémentaire à ce niveau. Le résultat net de ces réactions est la restauration de la séquence originale de l'ADN. Ce système sert vraisemblablement à protéger l'ADN contre les conséquences de la désamination spontanée de la cytosine (bien qu'il ne soit pas assez actif pour éviter les effets de l'acide nitreux ; voir Figure 2.17).

Au contraire, cependant, la désamination de la 5-méthylcytosine laisse une thymine ; puisque cette base est un constituant parfaitement respectable de l'ADN, le système de réparation n'opère pas dans ces circonstances et il en résulte une mutation.

Le fonctionnement de ce système jette une lumière intéressante sur l'utilisation de T dans l'ADN comparée à celle de U dans l'ARN. Elle s'associe peut-être au besoin de l'ADN d'avoir une séquence stable ; l'utilisation de T signifie que toutes les désaminations de C sont immédiatement reconnues parce qu'elles produisent une base (U) qui n'est pas habituellement présente dans l'ADN.

LE TAUX DE MUTATION

Les mutations spontanées qui inactivent un gène apparaissent chez les bactéries avec un taux d'environ 10^{-5} à 10^{-6} événements par locus et par génération. Nous n'avons pas de mesure réellement exacte du taux de mutation chez les eucaryotes, mais on pense qu'il est assez semblable à celui des bactéries exprimé par locus et par génération. Nous ne savons pas quelle est la proportion des événements spontanés due aux mutations ponctuelles.

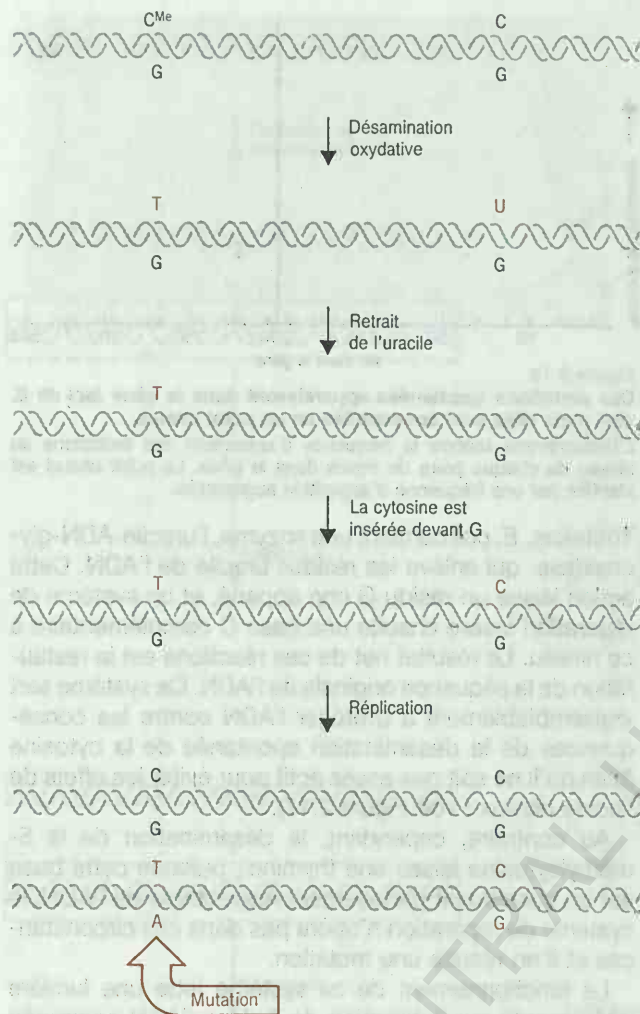


Figure 2.20

La désamination de la 5-méthylcytosine produit de la thymine (causant des transitions de C-G vers T-A) alors que la désamination de la cytosine produit de l'uracile (qui est habituellement retiré et remplacé par la cytosine).

Supposons qu'un gène bactérien soit constitué de 1 200 paires de bases codant pour une protéine de 400 acides aminés (d'une masse d'environ 45 000 daltons), le taux de mutation moyen correspond à 10^{-9} - 10^{-10} changements par génération au niveau de chaque nucléotide. Même si tous les changements résultent de mutations ponctuelles, ce calcul est extrêmement simplifié, car en fait toutes les mutations dans l'ADN ne conduisent pas à un changement détectable dans le phénotype.

Les mutations sans effet apparent sont appelées des **mutations muettes**. Il y en a deux types. Certaines

entraînent des changements de bases dans l'ADN qui ne provoquent aucun changement de l'acide aminé présent dans la protéine correspondante. D'autres remplacent l'acide aminé, mais le remplacement dans la protéine n'affecte pas l'activité de celle-ci ; ce sont des **substitutions neutres**.

Les mutations qui inactivent un gène sont des **mutations directes**. Leurs effets peuvent être réversés par des **mutations reverses**, qui sont de deux types.

Une inversion exacte de la mutation d'origine est dite **reversion vraie**. Ainsi, si une paire A-T a été remplacée par une paire G-C, une autre mutation qui restitue la paire A-T régènera exactement la séquence de type sauvage.

D'autre part, une autre mutation peut apparaître ailleurs sur le gène, et ses effets peuvent compenser ceux de la première mutation. C'est une **reversion à un second site**. Par exemple, le changement d'un acide aminé dans une protéine peut en supprimer la fonction, mais une seconde modification peut compenser la première et restituer l'activité de la protéine. (Ainsi les mutations compensatrices provoquées par les acridines par décalage du cadre de lecture appartiennent à la catégorie des révertants au second site.)

Une mutation directe est le résultat de toute modification qui inactive un gène, alors qu'une mutation réverse doit restituer la fonction à une protéine endommagée par une mutation directe spécifique. Ainsi les exigences pour une mutation reverse sont beaucoup plus spécifiques que celles pour une mutation directe. Le taux de mutation réverse est en conséquence plus bas que celui de mutation directe, d'un facteur d'environ 10.

POUR EN SAVOIR PLUS

Des souvenirs sur le développement de la biologie moléculaire ont été édités par Cairns, Stent & Watson, in *Phage and the Origins of Molecular Biology* (Cold Spring Harbor Laboratory, New York, 1966). Des comptes rendus historiques de cette période ont été écrits par Olby, *The Path to the Double Helix* (Mac Millan, London, 1974) et Judson, *The Eighth Day of Creation*. (Knopf, New York, 1978). Le cheminement pour caractériser les mutations a été revu par Drake & Balz (*Ann. Rev. Biochem.*, 45, 11-37, 1976). La base des mutations qui décalent le cadre de lecture a été revue par Roth (*Ann. Rev. Genet.*, 8, 319-346, 1974). Leur utilisation pour définir la nature du code génétique a été rapportée par Crick et al. (*Nature*, 192, 1227-1232, 1961). La découverte des points chauds a été rapportée par Benzer (*Proc. Nat. Acad. Sci. USA*, 47, 403-416, 1961) et (presque vingt ans plus tard) ils furent assimilés à des bases modifiées par Coulondre et al. (*Nature*, 274, 775-780, 1978).

CHAPITRE 3

LA TOPOLOGIE DES ACIDES NUCLÉIQUES

L'ADN est plus qu'une simple séquence de paires de bases organisées dans une structure en double hélice immuable. La grande majorité du génome est organisée selon le double-brin de forme B décrit au chapitre 2. Mais cette structure double est flexible et la capacité de modifier son organisation est un aspect fondamental de la fonction de l'ADN.

Le champ des changements possibles s'étend des altérations mineures des paramètres caractérisant le double-brin de forme B à la séparation des chaînes. Les caractéristiques d'importance physiologique sont :

- Le nombre de paires de bases par tour dans l'ADN-B n'est pas forcément définitif, mais peut être légèrement corrigé, selon les circonstances.
- La double hélice n'est pas une longue baguette droite mais elle s'enroule dans l'espace pour s'ajuster aux dimensions de la cellule ou du virus dont elle constitue le génome. Ce degré supplémentaire d'organisation pourrait placer le double-brin sous une contrainte telle, que sa propre structure en serait affectée.
- Certaines séquences peuvent avoir tendance à adopter une conformation double-brin autre que celle de l'ADN-B.
- Des discontinuités dans la structure peuvent prendre la forme de séquences spéciales où apparaissent des courbures.
- Pour qu'il y ait réplication ou transcription de l'ADN, les chaînes de la double hélice doivent être séparées. La séparation des brins n'est jamais que temporaire,

et, seules des régions relativement courtes du génome, apparaissent sous une forme simple-brin à un moment donné.

Bien que l'on trouve généralement l'ADN sous la forme d'une double hélice, les génomes de certains virus sont constitués d'un simple-brin d'ADN. A l'intérieur de la cellule existent plusieurs formes de l'autre acide nucléique, l'ARN, habituellement présent sous forme de simple-brin. Cependant les simples-brins d'ADN ou d'ARN peuvent engendrer des régions en double hélice, soit par appariement avec un simple-brin indépendant mais complémentaire, soit par un couplage de deux séquences complémentaires à l'intérieur d'une même molécule simple-brin.

Le concept de l'appariement des bases est au centre de tous les développements dans lesquels sont impliqués les acides nucléiques. La séparation des paires de bases est un aspect fondamental de la fonction d'une molécule double-brin, cependant que la capacité d'un acide nucléique simple-brin à former des paires de bases est essentielle à son activité.

L'ADN PEUT ÊTRE DÉNATURÉ ET RENATURÉ

Si l'on considère ses fonctions d'auto-réplication et de transcription en protéine, la propriété principale de la double hélice est la capacité qu'ont ses deux chaînes à se séparer sans avoir besoin de rompre les liaisons covalentes. Il est donc possible, pour les chaînes, de se

séparer et de se reformer sous certaines conditions physiologiques, et à une vitesse (très rapide) nécessaire au maintien des fonctions génétiques. La spécificité du processus est déterminée par l'appariement des bases complémentaires.

Ces caractéristiques mêmes, qui permettent à l'ADN de remplir son rôle biologique, permettent la manipulation de l'acide nucléique *in vitro*, et en dernier ressort, l'isolement du segment d'ADN qui représente une protéine donnée.

Les forces non-covalentes qui stabilisent la double hélice peuvent être brisées par une exposition à la chaleur ou dans un milieu fortement salin. Les deux chaînes d'une double hélice se séparent complètement lorsqu'elles ont rompu toutes les liaisons hydrogène.

Le processus de séparation des brins est appelé **dénaturation** (ou dans un langage plus courant) **fusion** (le terme « dénaturation » est également utilisé pour qualifier la perte de structure native d'une protéine ; c'est un terme général, impliquant que la structure naturelle d'une macromolécule a été changée en une autre forme).

La dénaturation de l'ADN se produit dans un champ restreint d'amplitude thermique, et il en résulte de très profonds changements affectant nombre de ses propriétés physiques. La modification de sa densité optique est particulièrement utile. Les hétérocycles des nucléotides absorbent fortement la lumière dans la zone des ultra-violet (avec un maximum proche de 260 mμ qui est caractéristique de chaque base). Mais l'absorption de l'ADN lui-même est de quelque 40 % inférieure à ce qu'elle serait pour un mélange de nucléotides libres de même composition.

On appelle cela l'effet **hypochrome** ; il résulte des interactions entre les systèmes d'électrons des bases, rendues possibles par leur entassement dans les rangées parallèles de la double hélice. Tout éloignement par rapport à l'état duplex se répercute immédiatement par une diminution de cet effet — c'est-à-dire par une augmentation de la densité optique vers la valeur caractéristique des bases libres. La dénaturation de l'ADN peut dès lors être suivie par cette **hyperchromicité**.

Le point médian de l'amplitude thermique auquel se produit la séparation des chaînes de l'ADN est appelé « **température de fusion** » et s'écrit T_m . Un exemple de courbe de fusion déterminée par un changement de la densité optique est montré en **Figure 3.1**. La courbe prend toujours la même forme, mais sa position absolue sur l'échelle de température (c'est-à-dire sa T_m) est influencée à la fois par la composition des bases de l'ADN et par les conditions utilisées pour la dénaturation.

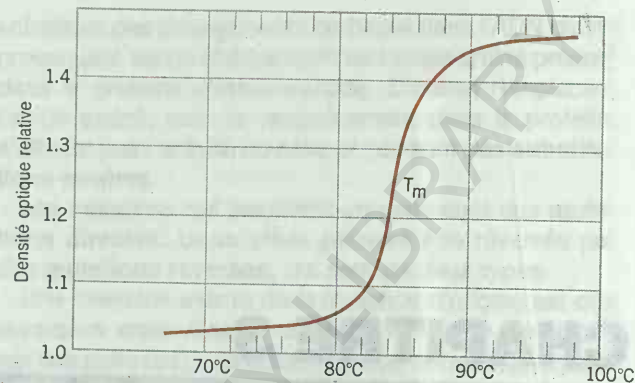


Figure 3.1

La dénaturation de l'ADN peut être suivie par l'augmentation de la densité optique et est définie par la T_m .

Lorsque l'ADN est en solution dans des conditions approximativement physiologiques, la T_m se situe habituellement entre 85-95 °C. Ainsi, sans l'intervention des systèmes cellulaires, le double-brin d'ADN est stable à la température prédominante dans la cellule. La **Note 3.1** résume quelques-uns des facteurs qui influencent la T_m .

LES ACIDES NUCLÉIQUES S'HYBRIDENT PAR APPARIEMENT DES BASES

Dans des conditions appropriées, la dénaturation de l'ADN est réversible. La capacité des deux chaînes complémentaires séparées, à se reformer en une double hélice, est appelée **renaturation** : elle est illustrée en **Figure 3.2**.

La renaturation dépend d'un appariement spécifique intervenant entre les chaînes complémentaires. La réaction s'effectue en deux phases. D'abord les simple-brins d'ADN se rencontrent par hasard dans la solution ; si leurs séquences sont complémentaires les deux brins se lient en formant une courte région en double hélice. Ensuite la zone de liaison s'étend le long de la molécule par un effet semblable à celui d'une fermeture à glissière, elle forme une double molécule allongée. La renaturation de la double hélice restitue les propriétés originelles perdues lors de la dénaturation de l'ADN.

Le terme renaturation définit la réaction entre deux séquences complémentaires qui ont été séparées par dénaturation. En outre, cette technique peut être développée et permettre à deux séquences d'acides nucléiques complémentaires, quelles qu'elles soient, de se **réassocier** l'une avec l'autre pour former une structure double-

Note 3.1

La T_m est déterminée par la composition des bases de l'ADN et par les conditions de dénaturation.

La valeur exacte de la T_m dépend de la proportion de paires de bases G-C ; chaque paire de bases G-C possédant trois liaisons hydrogène, elle est plus stable qu'une base A-T qui en possède seulement deux. Plus il y a de paires de bases G-C dans l'ADN, plus l'énergie requise pour séparer les deux chaînes doit être importante.

La relation entre la T_m et la composition en bases est linéaire, elle s'élève d'environ 0,4 °C pour chaque augmentation de 1 % de sa teneur en G-C. Ainsi un ADN qui contient 40 % de G-C (valeur typique des génomes mammifères) est dénaturé à une T_m d'environ 87 °C dans des conditions normales, alors qu'un ADN qui contient 60 % de G-C a une T_m de 95 °C dans les mêmes conditions.

La force ionique de la solution exerce un effet majeur sur la T_m . La T_m augmente de 16,6 °C chaque fois que décuple la concentration en cations monovalents. La condition la plus courante pour manipuler l'ADN *in vitro* est d'utiliser un tampon phosphate 0,12 M, à la concentration usuelle de 0,18 M en cation Na^+ , la T_m est voisine de 90 °C.

La T_m peut varier considérablement si l'on réalise la réaction en présence de réactifs, telle le formamide, qui déstabilise les liaisons hydrogène. Cela permet d'abaisser la T_m jusqu'à 40 °C, avec l'avantage pour l'ADN de ne subir aucun des dommages (telle la rupture de chaîne) qui pourraient résulter d'une exposition à des températures élevées.

brin. La réaction est généralement appelée **hybridation**, quand des acides nucléiques de sources différentes sont employés, comme c'est le cas lorsqu'une préparation est composée d'ADN et une autre d'ARN. La capacité de deux préparations d'acide nucléique à s'hybrider constitue une indication précise de la complémentarité de leurs séquences.

Le principe de la réaction d'hybridation est de mettre en contact deux préparations d'acide nucléique simple-brin et ensuite de mesurer la quantité de matériel double-brin formé. Il existe deux façons de réaliser la réaction : **L'hybridation en solution (liquide) et l'hybridation sur filtre.**

L'hybridation liquide est définie par son nom : les deux préparations d'ADN simple-brin sont mélangées en solution. Quand on emploie une importante quantité de matériel, la réaction peut être suivie grâce au changement de densité optique. Si les quantités de matériel sont moindres une des deux préparations pourra être marquée radioactivement, son appariement avec le brin complémentaire (c'est-à-dire sa renaturation) étant suivi par l'appariement de double-brin marqué. L'ADN double-brin peut être analysé, soit directement en utilisant la chromatographie pour séparer les double-brins des mono-brins, soit en dégradant tous les simple-brins n'ayant pas réagi et en mesurant ensuite la quantité de matériel restant.

L'hybridation en solution n'est pas une technique adaptée à l'étude des relations entre deux préparations dont l'une (ou les deux) est (sont) composée(s) d'ADN double-brin. Le problème est que si deux préparations d'ADN double-brin sont dénaturées et qu'ensuite les

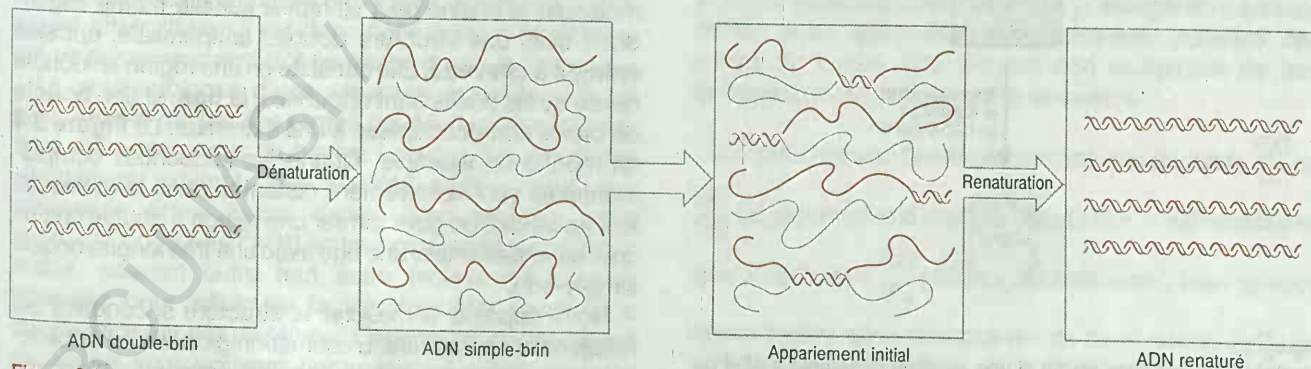


Figure 3.2

Des simple-brins d'ADN dénaturé peuvent être renaturés sous forme de double-brins.

mono-brins sont mélangés, il se passe deux types de réaction : les *simple-brins originaux* complémentaires peuvent se renaturer, ou bien chaque simple-brin peut s'hybrider avec une séquence complémentaire dans l'autre ADN. La compétition entre les deux réactions rend difficile l'estimation de l'ampleur de l'hybridation.

Cette difficulté peut être surmontée si l'on immobilise une des préparations d'ADN de telle façon qu'elle ne puisse pas se renaturer. Les filtres de nitrocellulose possèdent la propriété utile d'adsorber les acides nucléiques simple-brins ; et si un filtre a été utilisé pour fixer de l'ADN, certains traitements pourront éviter la fixation de tout autre mono-brin.

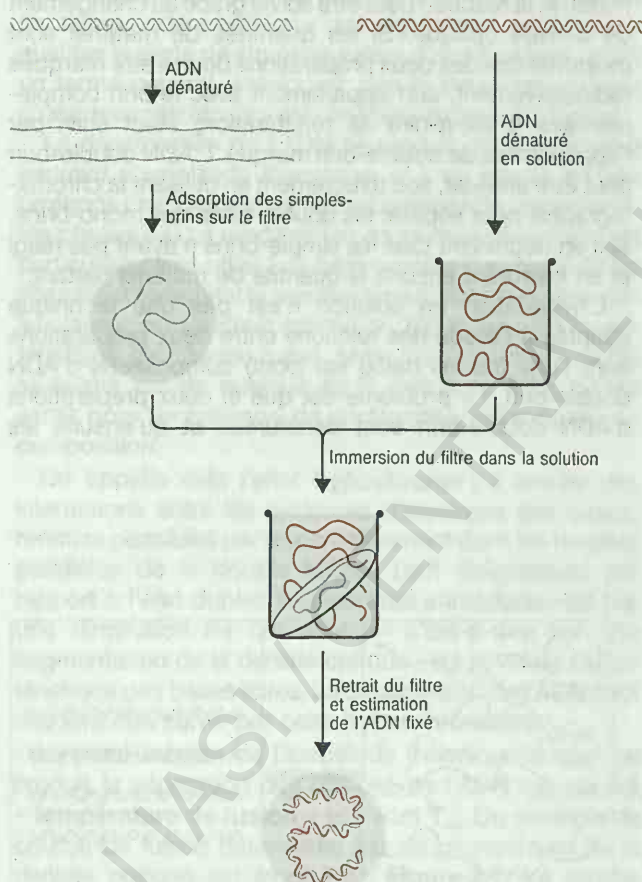


Figure 3.3
L'hybridation sur filtre établit si une solution dénaturée d'ADN ou d'ARN contient des séquences complémentaires aux brins immobilisés sur le filtre.

La **Figure 3.3** montre la méthode selon laquelle la préparation d'ADN est dénaturée et les simples-brins adsorbés sur le filtre. On ajoute ensuite une seconde préparation d'ADN dénaturé (ou d'ARN). Celle-ci n'est adsorbée sur le filtre que si elle est capable de s'apparier avec l'ADN déjà fixé. La procédure expérimentale courante consiste à ajouter au filtre une préparation d'ADN ou d'ARN radioactif, ce qui permet d'évaluer la réaction en mesurant la quantité de radioactivité retenue par le filtre.

LES ACIDES NUCLÉIQUES MONO-CATÉNAIRES PEUVENT AVOIR UNE STRUCTURE SECONDAIRE

La stabilité de la double hélice dépend des liaisons hydrogène entre les paires complémentaires A-T et G-C, et également des interactions entre les bases puisqu'elles sont « empilées » les unes sur les autres le long de l'axe de l'hélice. Ces forces peuvent être utilisées pour prévoir la stabilité d'une double hélice entre deux séquences complémentaires. De telles extrapolations sont utilisées pour analyser la structure d'une séquence d'acide nucléique mono-caténaire. L'ARN étant l'acide nucléique mono-brin majoritaire, la technique de formation de double-brins à partir de simple-brin est surtout utilisée pour l'analyse de l'ARN mais elle s'applique également à l'ADN simple-brin.

La structure primaire de l'ARN est identique à celle de l'ADN : une chaîne de polynucléotides avec des liaisons sucre-phosphate 5'-3'. Considérée en tant que simple-brin, la molécule suit dans l'espace un cheminement au hasard, mais des appariements intra-moléculaires peuvent fixer l'emplacement d'une région apparentée à une autre.

Lorsqu'une séquence de bases est suivie d'une séquence complémentaire toute proche dans la même molécule, la chaîne peut se replier sur elle-même engendrant ainsi une structure double, antiparallèle, appelée **épingle à cheveux**. Elle consiste en une région en double hélice où les bases sont appariées, la **tige**, et une **boucle** de bases non-accouplées à une extrémité. La **Figure 3.4** en montre un exemple. Quand les séquences complémentaires sont relativement distantes dans la molécule, leur juxtaposition pour former une région à double chaîne crée essentiellement une tige avec une très longue boucle simple-brin.

Notre capacité à mesurer la structure secondaire est relativement sommaire. L'estimation *globale* de l'appariement des bases se reflète dans les propriétés biophysiques de la molécule. Toutefois cela ne nous révèle pas

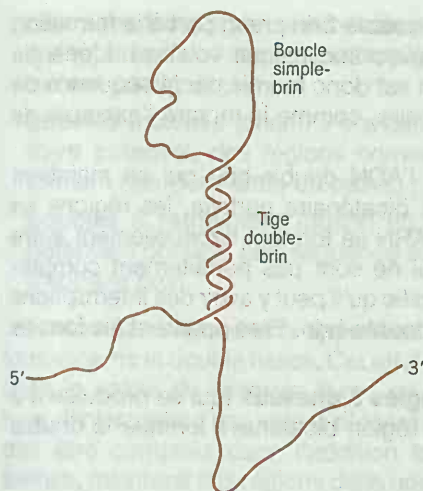


Figure 3.4

Un acide nucléique mono-caténaire peut se replier sur lui-même pour former une épingle à cheveux par appariement des bases de séquences complémentaires.

quelles régions particulières sont concernées. Les régions simple-brins et les régions double-brins ont des sensibilités différentes à certaines **nucléases** (enzymes qui dégradent les acides nucléiques) et cela permet d'analyser l'implication de régions spécifiques dans l'appariement des bases. En général cependant, de telles données ne sont indicatives qu'avec des molécules relativement courtes.

Jusqu'à quel point peut-on prévoir la présence d'une structure secondaire dans les séquences d'acide nucléique simple-brins ? Il ne s'agit pas seulement de décider simplement si les structures secondaires sont à même d'exister ou non, car une molécule simple-brin peut posséder plusieurs régions potentiellement capables de former des paires de bases entre elles, selon des dispositions alternatives ; il est donc nécessaire d'élucider ce qui se passe vraiment (s'il se passe quelque chose).

La *vraisemblance* d'une structure appariée donnée peut être prévue par des **règles** qui décrivent les interactions entre les paires de bases. Lorsque des structures alternatives existent, leur stabilité relative peut être démontrée grâce à ces règles. Bien entendu, cette approche considère l'ARN en tant que structure isolée et stable, passant outre tout autre facteur qui pourrait intervenir pour influencer la structure (par exemple la stabilisation par des protéines). Cependant, en appliquant ces règles nous pouvons commencer à entrevoir si une structure donnée peut, ou non, exister.

La base de ces règles est de calculer l'**énergie libre** nécessaire à la formation de chaque structure. L'énergie libre est (en résumé) une constante thermodynamique qui donne la quantité d'énergie libérée lors d'une réaction. Elle se mesure en kcal/mole et s'exprime sous la forme du paramètre ΔG qui a une valeur *négative* correspondant à l'énergie disponible. Ainsi si des structures alternatives ont des énergies libres de formation de $\Delta G = -21$ kcal/mole et $\Delta G = -35$ kcal/mole, il est plus probable que cette dernière se produise, toutes choses étant égales par ailleurs.

L'énergie libre globale d'une structure à double chaîne s'obtient en additionnant les énergies libres des réactions individuelles d'appariement des bases qui sont impliquées. La formation de chaque liaison hydrogène ne dégage qu'une quantité relativement faible d'énergie. Comme les paires de bases G-C ont trois liaisons hydrogène, elles sont les plus stables, avec un ΔG de $-2,4$ kcal/mole ; les deux liaisons hydrogène d'une paire A-U donnent un ΔG de $-1,2$ kcal/mole. Cela nous donne une base quantitative pour l'exposé de la Note 3.1 qui montre que la stabilité d'une double hélice augmente avec la proportion des paires de bases G-C.

Une autre paire de bases qui peut se former est l'association illégitime G-U (et G-T qui peut se former dans des régions double-brin d'ADN simple-brin bien qu'elle ne fasse pas partie de la double hélice normale). Cela donne un ΔG de 0 ; en d'autres termes, un G peut faire face à un U dans une structure d'ARN en double hélice sans rompre l'hélice ni contribuer à sa formation.

L'énergie libre est produite par les interactions hydrophobes existant entre les bases, ces dernières étant empilées l'une sur l'autre à l'intérieur de l'axe hélicoïdal (l'eau est exclue de l'intérieur de l'hélice ce qui permet ainsi un environnement hydrophobe). L'énergie libre de l'empilage des bases dépend de la combinaison particulière des bases azotées contiguës ; l'énergie libre globale diffère donc pour chaque combinaison possible des paires de bases. Elle dépend non seulement de leur composition mais aussi de la séquence.

On peut donner l'exemple suivant, les doublets $\begin{smallmatrix} \text{GG} \\ \text{CC} \end{smallmatrix}$ et $\begin{smallmatrix} \text{GC} \\ \text{CG} \end{smallmatrix}$ dégagent plus d'énergie libre ($\Delta G = -5,0$ kcal/mole)

que la séquence $\begin{smallmatrix} \text{CG} \\ \text{GC} \end{smallmatrix}$ ($\Delta G = -3,2$ kcal/mole), bien qu'elles soient toutes deux composées de deux paires G-C. De même, la séquence $\begin{smallmatrix} \text{AA} \\ \text{UU} \end{smallmatrix}$ dégage moins d'énergie ($\Delta G =$

hélice est formée par une boucle en retour, de façon à s'apparier à une séquence complémentaire immédiatement contiguë) ;

boucles internes (quand il y a dans chaque complément potentiel, des régions correspondantes qui ne forment pas des paires de bases) ;

boucles à renflement (quand un des compléments potentiels contient une séquence surnuméraire qui ne s'apparie pas).

Toutes ces régions non appariées *font obstacle* à la formation de la double hélice. Cet effet est pris en compte dans le calcul de l'énergie libre : on attribue à chaque type d'interruption une valeur d'énergie libre *positive* qui doit être comprise dans l'addition totale. En d'autres termes, maintenir ces régions dans une structure particulière requiert « *un apport* » d'énergie, aux dépens de l'énergie libre dégagée par l'accouplement des bases et leur empilage. Le calcul de l'énergie libre doit produire une valeur globale négative significative, sinon la structure secondaire ne pourra pas se former.

RÉPÉTITIONS INVERSES ET STRUCTURE SECONDAIRE

Lorsqu'un acide nucléique monocaténaire contient deux séquences complémentaires, elles peuvent s'apparier pour constituer une épingle à cheveux comme il est montré en Figure 3.4. La séquence, qui, dans une double chaîne d'ADN, correspond à une telle structure consiste en deux copies d'une séquence identique présentes en orientation inverse. Elles sont appelées **répétitions inverses**.

La séquence globale des répétitions inverses est appelée un **palindrome**. C'est une séquence d'ADN double qui se révèle identique lorsque l'on lit chacun de ses brins dans une direction donnée.

5'	GGTACC	3'
3'	CCATGG	5'

est un exemple de palindrome car lorsque l'on lit chacun des brins dans l'orientation 5' vers 3', cela nous donne la séquence GGTACC (c'est-à-dire de gauche à droite pour la rangée supérieure et de droite à gauche pour la rangée inférieure).

Formellement une séquence palindromique est définie comme une **région de symétrie** par rapport à un point (dyad symmetry), le **centre de symétrie** est le point

central à partir duquel la séquence est la même de chaque côté (lorsqu'on lit dans des directions opposées). En traçant une ligne passant par le centre de symétrie

GGT	ACC
CCA	TGG

on voit que la même séquence **GGT CCA** est présente de part et d'autre de cette ligne, dans des orientations inverses, comme l'indiquent les flèches. Chaque flèche représente une des répétitions inverses.

Les deux copies d'une répétition inverse n'ont pas nécessairement besoin d'être contiguës. Considérons la même répétition inverse dans une situation différente

GGTNN	NNACC
CCANN	NNTGG

Il y a toujours une répétition inverse du même triplet, mais le centre de symétrie a été déplacé au centre des quatre paires de bases supplémentaires qui séparent les deux copies de la répétition.

On peut trouver une répétition inverse dans la structure secondaire de chaque acide nucléique, mono ou bicaténaire. Si l'on considère un seul brin, les copies de la répétition sont complémentaires quand on les lit dans des directions opposées. Lorsque l'on lit à partir du centre de symétrie, on obtient TGG vers la gauche et ACC vers la droite. La situation est semblable à celle illustrée en Figure 3.5 et également en **Figure 3.6** ; les séquences complémentaires peuvent s'unir pour ensuite produire une épingle à cheveux.

Dans une double chaîne d'ADN, les séquences complémentaires présentes sur un brin ne peuvent s'apparier que si ce brin se sépare de son partenaire. La situation est la même sur chaque brin (ou chaîne) : une épingle à cheveux peut se former. La formation de deux épingles à cheveux accolées crée une **structure cruciforme**, que l'on appelle ainsi parce qu'elle marque la jonction de quatre régions double-brins. Le double-brin d'origine s'étend de chaque côté, et les épingles à cheveux intra-fibres en divergent.

Les conditions dans lesquelles les structures cruciformes peuvent se former *in vivo* sont controversées. Il est peu probable qu'une double hélice abandonne sa structure bicaténaire continue pour former spontanément des boucles en épingles à cheveux saillantes. Il faudrait une quantité considérable d'énergie pour séparer les deux brins (ΔG d'environ + 50 kcal/mole) et seule une petite quantité de cette énergie serait récupérée dans la formation de la structure cruciforme, laquelle est énergétique-

ADN cruciforme

Réarrangement

Double-brin d'ADN

Transcription

ARN en épingle à cheveux

Répétitions inversées

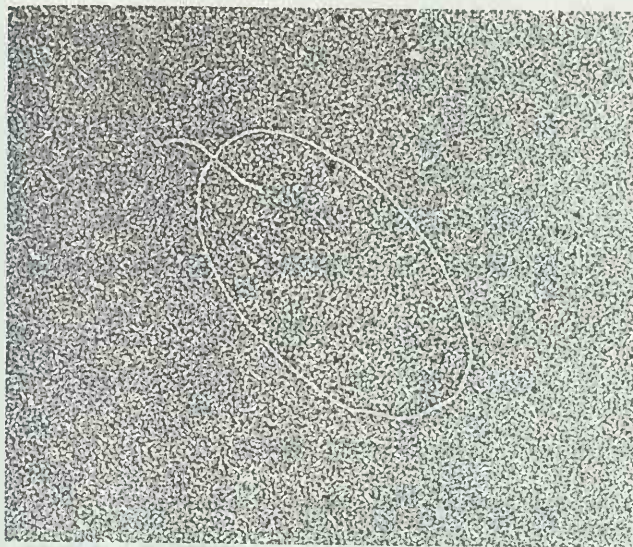
Figure 3.6

Les répétitions inversées peuvent engendrer des structures cruciformes dans l'ADN ou des épingles à cheveux dans l'ARN.

Le double-brin d'ADN au centre montre les séquences de répétitions inversées. Une structure cruciforme serait engendrée si un réarrangement amenait chaque répétition à s'hybrider à la séquence complémentaire de son propre brin et non au complément de l'autre brin.

La transcription du brin supérieur d'ADN engendrerait un ARN, de séquence semblable à celle du brin inférieur d'ADN. Les répétitions peuvent s'apparier pour engendrer une épingle à cheveux qui a la même séquence que celle de la partie inférieure de la structure cruciforme.

Les répétitions inversées forment elles-mêmes la tige de double-brins de l'épingle à cheveux. Les trois bases qui séparent les répétitions inversées apparaissent sous la forme d'une boucle simple-brin à l'extrémité de la tige.

**Figure 3.7**

Photographie en microscopie électronique d'une structure cruciforme dans de l'ADN *in vitro*.

Photographie gracieusement fournie par Martin Gellert.

ment moins stable qu'un double-brin continu (ΔG entre les deux structures étant de + 18 kcal/mole). Cependant des structures cruciformes peuvent être engendrées si les conditions appropriées sont réunies *in vitro*, comme on peut le voir sur l'exemple de la **Figure 3.7**.

Nous ne savons pas si de telles structures peuvent apparaître *in vivo*. L'opinion la plus répandue tendrait à dire que les structures cruciformes ne se forment probablement pas naturellement dans les cellules, et si l'on peut les faire apparaître dans de l'ADN *in vitro* à l'aide de séquences appropriées, elles seront probablement instables si on les introduit dans une cellule.

L'ADN DOUBLE-BRIN A DES STRUCTURES EN DOUBLE-HÉLICE ALTERNATIVES

Cela fait longtemps que l'on sait que l'ADN peut exister dans plus d'un type de structure en double-hélice, bien que dans les conditions physiologiques, il apparaisse toujours sous la forme désormais classique du double-brin de forme B. De plus, nous pensons toujours que le double-brin d'ADN est presque entièrement sous la forme B à l'intérieur de la cellule, bien que certains des paramètres définissant la structure précise du double-brin de type B aient été affinés, en particulier, le nombre de paires

de bases par tour. Quoi qu'il en soit, dans certaines conditions, le double-brin d'ADN peut être capable d'opérer une transition de la forme B vers une autre forme. Cela affecte probablement une proportion relativement réduite de l'ADN.

La mesure du nombre de paires de bases par tour dans une double-chaîne d'ADN donne un résultat de 10,4 au lieu de 10,0 prévu par le modèle classique B. Le changement requiert un léger ajustement dans l'angle de rotation entre les paires de bases contiguës le long de l'hélice, à $34,6^\circ$, de sorte qu'il faudra légèrement plus de temps pour accomplir le tour complet à 360° .

Le double-brin B nous fournit un modèle pour déterminer ce que l'on pourrait appeler une donnée « moyenne » ; des variations peuvent se produire dans certaines régions ; elles résultent soit d'une séquence de bases spécifiques, ou des conditions imposées par l'environnement. Nous savons que la structure précise peut effectivement être altérée ; par exemple, on a montré par cristallographie aux rayons X que la structure d'une molécule spécifique de 12 paires de bases avait 10,1 paires de bases par tour, chaque paire de bases accomplissant une légère torsion qui améliore l'empilage des bases par rapport au modèle d'origine.

In vivo, l'enroulement serré et le compactage, qui sont nécessaires pour que l'ADN tienne dans la cellule, peuvent changer des détails de sa structure. Ainsi, nous prenons la valeur de 10,4 paires de bases par tour pour représenter une moyenne pour un ADN dans son ensemble, dans certaines conditions, et nous supposons que la valeur de chaque séquence spécifique sera proche de cette estimation, mais pas nécessairement identique.

L'idée que la double-hélice d'ADN possède une structure unique, a été remplacée par l'idée qu'il existe des familles de structures. Chaque famille représente un type caractéristique de double-hélice, comme le démontrent les paramètres n (nombre de nucléotides par tour) et h (distance entre les unités répétées contiguës). La variation s'accomplit par des modifications dans la rotation des groupes autour des ponts ayant une liberté de rotation. Dans chaque famille les paramètres peuvent varier légèrement ; par exemple, pour un ADN-B, n peut varier entre 10,0 et 10,6.

Le **Tableau 3.1** résume les propriétés moyennes de quatre familles de structure. On connaît depuis longtemps l'existence des formes A, B et C, des formes intermédiaires se produisant lorsque les conditions sont modifiées. Ce sont toutes des structures tournant à droite (ou droites) dont les paramètres sont légèrement différents. Toutefois, la forme Z représente une structure relativement différente qui a été découverte plus récemment.

Tableau 3.1

L'ADN peut se présenter sous plusieurs types de structures.

Type d'hélice	Paires de bases/tour	Rotation par paire de bases	Élévation verticale par paire de bases	Diamètre hélicoïdal	Conditions d'apparition	
					En fibres	En solution
A	11	+32,7°	2,56 Å	23 Å	75 % humidité	K ⁺ Na ⁺ ou Cs ⁺
B	10	+36,0°	3,38 Å	19 Å	92 % humidité	faible force ionique
C	9,33	+38,6°	3,32 Å	19 Å	66 % humidité	—
Z	12	-30,0°	3,71 Å	18 Å	—	force ionique élevée

La rotation par paire de bases est indiquée par un (+) pour un double-brin droit et par un (-) pour un double-brin gauche.

La forme C n'existe probablement pas *in vivo*. Il existe deux autres formes, la **forme D** et la **forme E** qui peuvent être des variantes extrêmes de la même forme ; elles ont le plus petit nombre de paires de bases par tour (seulement 8 et 7 1/2) et sont prises *in vitro* par certaines molécules d'ADN manquant de guanine.

La cellule vivante fournit les conditions appropriées pour la formation de la **forme B**, forme B d'après laquelle Watson et Crick ont construit leur modèle, et qui représente la structure générale de l'ADN.

La **forme A** est probablement très proche de la configuration des régions d'ARN à doubles-chaînes, où la présence du groupe 2'-hydroxyle rend impossible la configuration de forme B. Les double-brins hybrides possédant un brin d'ADN et un brin d'ARN existent probablement sous la forme A.

La **Figure 3.8** compare les formes courantes des acides nucléiques double-brins. Il est probable que les motifs structuraux nécessaires à la reconnaissance d'un double-brin soient différents dans un ADN de forme B et dans un ARN de forme A.

L'ADN-B a un sillon important ou grand sillon, et un sillon mineur ou petit sillon. Les différences entre les bases sont plus facilement identifiables dans le grand sillon, qui est un point de contact privilégié pour les protéines se fixant sur des séquences d'ADN spécifiques.

Dans la structure relativement compacte de l'ARN de forme A, les bases au lieu de reposer à plat sont inclinées par rapport à l'axe hélicoïdal, et le nombre de paires de bases par tour est augmenté. Le grand sillon est beaucoup moins accessible car il est situé plus en profondeur et il est surmonté de groupements phosphate. Les caractéristiques des bases individuelles sont donc enterrées profondément dans la rainure. Le petit sillon cependant est superficiel.

UNE FORME D'ADN TOURNANT À GAUCHE

La **forme Z** nous fournit le contraste le plus frappant avec les familles de structures classiques. C'est la seule **hélice tournant à gauche**. Sa structure est illustrée en **Figure 3.8**.

L'ADN-Z possède plus de paires de bases par tour que toutes les autres formes double-brins, et elle a donc la structure la moins enroulée ; elle est éflanquée, et son nom est tiré du trajet en zig-zag que suit le squelette sucre-phosphate le long de l'hélice, celui-ci très différent du trajet en lacets réguliers que suit le squelette d'un ADN de forme B. L'ADN-Z ne possède qu'un seul sillon, mais d'une densité en charges négatives supérieure à celle de chacun des deux sillons de l'ADN-B.

La double hélice de forme Z apparaît dans des polymères dont la séquence est composée de purines et de pyrimidines en alternance. Deux exemples étudiés montrent une séquence simple constituée de dinucléotides répétés :

poly d $\begin{Bmatrix} \text{GC} \\ \text{CG} \end{Bmatrix}$ et poly d $\begin{Bmatrix} \text{AC} \\ \text{TG} \end{Bmatrix}$. (Le *d* indique que ce sont des formes déoxy et une séquence d'ADN et non d'ARN).

L'ADN de forme Z a été découvert *in vitro* dans des conditions quelque peu inhabituelles (en utilisant une concentration salée très élevée pour contrecarrer la répulsion électrostatique accrue entre les nucléotides comprimés dans la mince double hélice d'ADN-Z). Dans quelles conditions la forme Z pourrait-elle exister *in vivo* ?

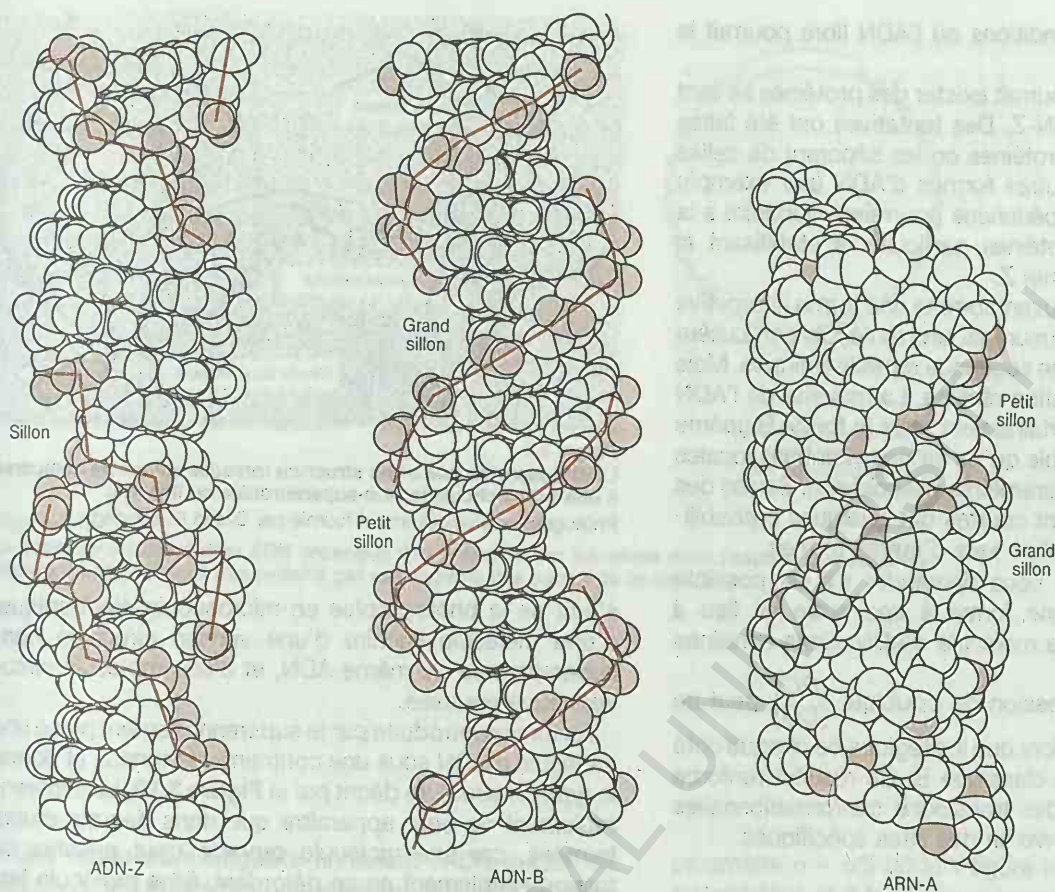


Figure 3.8
Les acides nucléiques peuvent former différents types de double hélice.

Deux facteurs intrinsèques de la double hélice déterminent la probabilité pour qu'une région spécifique d'ADN existe dans la forme Z. L'un est la séquence nucléotidique ; et l'autre est la structure globale de la double hélice selon son architecture dans l'espace, un effet que nous montrerons dans la prochaine partie du livre. Si ces deux facteurs sont propices, l'ADN peut être capable de se convertir de la forme B à la forme Z dans les conditions naturelles.

L'exigence la plus importante est celle d'une alternance de purines et de pyrimidines mais la nature des bases elles-mêmes exerce également une influence. En particulier, remplacer le résidu C dans le polymère poly d $\begin{Bmatrix} \text{GC} \\ \text{CG} \end{Bmatrix}_n$ par une 5-méthylcytosine rend l'ADN-Z beaucoup plus stable à des concentrations salines moins élevées.

La différence entre 5-Me-C et C tient à la présence du groupe méthyle en position 5. Cette modification de la cytosine a lieu (en certains endroits) *in vivo*, dans des séquences dinucléotidiques $\begin{Bmatrix} \text{CG} \\ \text{GC} \end{Bmatrix}$. Cela rend plus plausible l'existence de l'ADN-Z *in vivo* si de bonnes conditions sont réunies, c'est-à-dire si une séquence, à la composition en bases appropriée, se trouve dans des conditions favorisant la formation de l'ADN-Z.

Jusqu'à présent, nous n'avons évoqué l'ADN que comme un acide nucléique dans son splendide isolement. En réalité, il est associé à des protéines qui peuvent affecter de manière substantielle sa capacité à opérer le glissement entre la forme B et la forme Z. Par exemple, l'ADN associé aux histones (les protéines basiques de la chromatine des noyaux eucaryotiques) ne peut opérer de

transition dans les conditions où l'ADN libre pourrait le faire.

D'un autre côté, il pourrait exister des protéines se liant spécifiquement à l'ADN-Z. Des tentatives ont été faites pour isoler de telles protéines en les séparant de celles qui s'associent aux autres formes d'ADN (par exemple l'ADN-B). De telles expériences pourraient conduire à la caractérisation de protéines particulières stabilisant *in vivo* l'ADN dans la forme Z.

L'ADN-Z a été découvert comme une forme alternative prise sur toute sa longueur, par une molécule particulière d'ADN composée d'une séquence répétitive simple. Mais ce n'est pas une situation réaliste. La majorité de l'ADN cellulaire se trouve certainement sous la forme B (même s'il est toutefois possible qu'il y ait des variations locales dans les valeurs des paramètres hélicoïdaux). Seules des séquences relativement courtes ont quelques probabilités de subir une transition vers d'autres formes.

Nous devons donc nous demander s'il est possible qu'un glissement d'une forme à une autre ait lieu à l'intérieur d'une même molécule d'ADN. Dans certaines

conditions, une succession de doublets $\begin{Bmatrix} GC \\ CG \end{Bmatrix}$ peut se convertir en ADN-Z, alors que les régions de chaque côté restent sous la forme classique B. Ce résultat renforce l'idée selon laquelle des transitions conformationnelles peuvent avoir lieu *in vivo* en des sites spécifiques.

L'ADN FERMÉ PEUT ÊTRE SUPERENROULÉ

La double hélice représente l'ADN comme une molécule linéaire. Mais généralement l'ADN *in vivo* a une structure **fermée** : il ne possède pas d'extrémités libres. Les génomes de certains petits virus sont, de fait, constitués d'**ADN circulaire**, dans lequel les deux brins de la double hélice parcourent de façon continue le tour du cercle. Dans les génomes bactériens et eucaryotes, l'ADN existe sous la forme de boucles plutôt larges ; les boucles sont reliées entre elles à la base, formant un simulacre de cercle. Cette forme d'organisation est importante parce qu'elle permet d'imposer à la double hélice une contrainte structurale supplémentaire.

Des **supertours** apparaissent dans l'ADN quand un double-brin subit une torsion dans l'espace autour de son propre axe. On a l'habitude de les comparer à un élastique qui serait enroulé sur lui-même pour former une structure torsadée dans laquelle l'élastique (la double hélice) se surcroise dans l'espace. Les formes possibles d'une molécule d'ADN sont comparées **Figure 3.9**, il

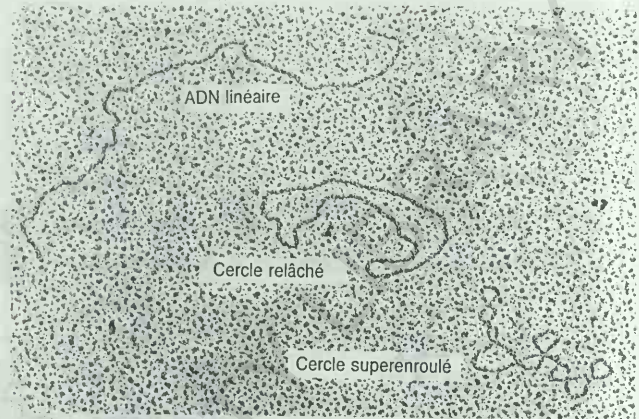


Figure 3.9

L'ADN superenroulé a une structure torsadée compacte comparée à une forme circulaire non-superenroulée ou linéaire.

Photographie gracieusement fournie par Svend O. Freytag.

s'agit de la photographie en microscopie électronique, d'une molécule linéaire d'une version circulaire non-superenroulée du même ADN, et d'une molécule circulaire superenroulée.

La torsion introduite par le superenroulement place une molécule d'ADN sous une contrainte de torsion et donne le type de structure décrit par la **Figure 3.10**. Le superenroulement ne peut apparaître que dans des structures fermées, car une molécule **ouverte** peut relâcher sa torsion simplement en se détordant. (Une molécule **fermée** ne doit avoir aucune coupure sur *aucune* des fibres d'ADN ; une coupure même sur un seul brin d'une molécule circulaire permet la détorsion.) Une molécule qui ne possède pas de superenroulement, qu'elle soit ouverte ou fermée, est dite **relâchée**.

Le principe à la base du superenroulement est que *lorsqu'un double-brin est enroulé sur lui-même dans l'espace, ses extrémités étant fixées, la structure ainsi constituée est incapable de se démêler d'elle-même*. Un supertour (ou plus formellement un tour de superenroulement) se produit chaque fois que le filament du double-brin est enroulé autour de son axe, et plus le nombre de supertours est grand, plus la torsion est importante dans une molécule fermée.

Pour un élastique, peu importe la *direction* dans laquelle on applique la torsion qui produit le superenroulement (les deux bords de l'élastique sont semblables). Mais la double hélice étant elle-même une structure enroulée (comme on le voit dans l'entrelacement des deux fibres) sa réponse à la torsion dépend de la direction du superenroulement.

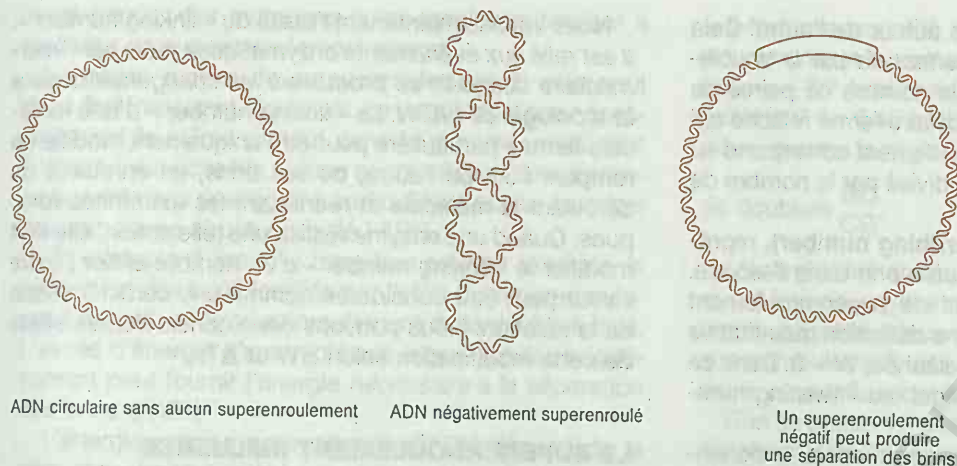


Figure 3.10

Le superenroulement d'un ADN provoque l'enroulement sur lui-même dans l'espace ; un superenroulement négatif peut être relâché par une rupture des paires de bases.

Les supertours négatifs enroulent l'ADN autour de son axe dans la direction *opposée* au sens des aiguilles d'une montre qui caractérise la double hélice droite. Cela permet en principe à l'ADN de relâcher la pression de torsion, en ajustant la structure de la double hélice elle-même. Le relâchement se traduit par une réduction de la rotation par paire de bases, c'est-à-dire qu'il y a perte de l'enroulement des deux brins entre eux. On dit de l'ADN qui possède des supertours négatifs qu'il est **sous-enroulé**. Si la torsion est suffisamment importante, cela peut même conduire à une séparation partielle des paires de bases, comme il est montré à la Figure 3.10.

L'effet inverse se produit si l'ADN est superenroulé dans le *même* sens que l'enroulement intrinsèque de la double hélice. L'introduction de **supertours positifs** resserre la structure, en appliquant une pression de torsion pour enrouler la double hélice de façon encore plus serrée. On dit d'un ADN qui contient des supertours positifs qu'il est **sur-enroulé**. Cet état peut être créé par traitement *in vitro*, mais ne peut se produire naturellement.

Le surenroulement de l'ADN est exprimé par la **densité superhélicoïdale**, le nombre de tours de super-hélice par tour de double hélice. Pour décrire cela on utilise le

paramètre $\sigma = \tau/\beta$ où le τ figure le nombre de tours de super-hélice et β le nombre de tours du double-brin, qui est approximativement le nombre de paires de bases divisé par 10. Ainsi σ correspond au nombre de tours de super-hélice pour environ 10 paires de bases. Ce nombre est négatif pour le superenroulement négatif et positif pour le superenroulement positif.

Depuis la définition originelle du superenroulement, une nouvelle nomenclature a été introduite. Ce concept se base sur le **nombre d'entrecroisements (linking number)** qui désigne le nombre total de fois où les deux brins de la double hélice se croisent. Le « linking number » L est le nombre de révolutions qu'un brin réalise autour de l'autre lorsque l'ADN est considéré (hypothétiquement) comme reposant à plat sur un plan horizontal. La convention veut que l'on évalue le « linking number » L positivement pour chaque entrecroisement dans une double hélice droite. C'est nécessairement un nombre entier.

Le « linking number » (L) a deux composantes, l'enroulement et la torsion, définies par l'équation :

$$L = W + T$$

Le **nombre d'enroulements T , (twisting number)**, est une propriété de la structure de l'hélice elle-même, qui

représente la rotation d'une fibre autour de l'autre. Cela définit le nombre total de tours effectués par le double-brin. Ainsi, il est déterminé par le nombre de paires de bases par tour. Pour un ADN circulaire fermé relâché qui repose à plat dans un plan, l'enroulement correspond au nombre total de paires de bases divisé par le nombre de paires de bases par tour.

Le nombre de torsion, W , (*writhing number*), représente la rotation de l'axe du double-brin dans l'espace. Il correspond au concept intuitif de superenroulement mais n'a pas exactement la même définition quantitative ou mesure. Pour une molécule relâchée, $W = 0$. Dans ce cas, le « linking number » est égal au « twisting number ».

Le paramètre mesuré dans de nombreuses expériences est la variation du « linking number » ΔL . Il est donné par l'équation

$$\Delta L = \Delta W + \Delta T$$

L'équation établit que tout changement du nombre total de révolutions d'un brin d'ADN autour de l'autre peut être exprimé comme la somme des variations de la torsion de l'axe du double-brin dans l'espace (ΔW) et des variations de l'enroulement de la double hélice elle-même (ΔT).

Selon la terminologie antérieure, ΔW correspond à la variation du superenroulement (appelé τ), et ΔT représente le sous (ou le sur) enroulement de la double hélice d'ADN (appelé β), on peut donc écrire :

$$\alpha = \beta + \tau$$

où α est le chiffre d'enroulement topologique (topological winding number), équivalent à L .

Une diminution du « linking number », c'est-à-dire une variation $-\Delta L$, correspond à l'introduction, dans la molécule d'ADN, d'une combinaison variable de superenroulement négatif et/ou de sousenroulement de la double hélice. Une augmentation du « linking number », soit un changement égal à $+\Delta L$, correspond à une diminution du superenroulement négatif et/ou du sousenroulement.

Le point critique concernant l'utilisation du « linking number » est que ce paramètre est une propriété invariable de chaque molécule individuelle fermée d'ADN. Ce paramètre ne peut être modifié par aucune déformation, si ce n'est celles qui impliquent la rupture et la reconstruction des brins. Une molécule circulaire avec un « linking number » particulier l'exprimera selon différentes combinaisons de T et de W , mais ne pourra pas modifier leur somme tant que les brins ne sont pas rompus. (En fait la partition de L entre T et W empêche l'attribution de valeurs fixes pour les paramètres évoqués ci-dessus, dans le cas d'une molécule d'ADN en solution).

Nous voyons maintenant l'utilité du « linking number ». Il est relié aux événements enzymatiques réels par l'intermédiaire desquels se produisent les changements dans la topologie de l'ADN. Le « linking number » d'une molécule fermée particulière peut être uniquement modifié en rompant l'un, ou l'autre, de ses brins, en enroulant ou déroulant la molécule et réunissant les extrémités rompues. Quand une enzyme réalise une telle action, elle doit modifier le « linking number » d'un nombre entier ; cette valeur peut être considérée comme une caractéristique de la réaction. Nous pouvons donc considérer les effets de cette modification selon ΔW et ΔT .

LE SUPERENROULEMENT INFLUENCE LA STRUCTURE DE LA DOUBLE HÉLICE

Des modifications dans la structure de l'ADN ne se produisent pas spontanément *in vivo*, elles peuvent se produire dans un environnement qui place l'ADN sous diverses contraintes. Certaines de ces modifications sont favorisées par le superenroulement négatif.

L'introduction d'un superenroulement négatif demande de l'énergie ; la molécule superenroulée a davantage d'énergie grâce aux supertours qu'elle possède. On pourrait les considérer comme une réserve d'énergie. Aussi, lorsqu'un changement dans une structure d'ADN est sur le point de se produire, et qu'il demande de l'énergie, une moindre demande d'énergie sera nécessaire si l'ADN est négativement superenroulé. En pratique, cela signifie que le superenroulement négatif peut influencer l'équilibre d'une modification de structure ; ainsi l'ADN superenroulé peut subir des transitions structurales dans des conditions qui ne permettraient pas à l'ADN relâché de les réaliser.

Peut-on quantifier l'effet d'un superenroulement négatif ? Tous les génomes qui ont été étudiés montrent une certaine quantité de superenroulement négatif. Le degré caractéristique *in vivo* est d'environ 1 tour négatif pour 200 paires de bases, ce qui correspond à une densité superhélicoïdale σ de $-0,05$. Cela confère à la molécule une énergie de quelque -9 kcal/mole.

Le superenroulement *in vivo* est contrôlé par un équilibre entre les activités des enzymes qui introduisent les supertours dans l'ADN et des enzymes qui les font disparaître (voir Chapitre 28). On a trouvé, chez la bactérie, des mutations qui affectent les activités de ces enzymes. Les souches qui possèdent des degrés altérés de superenroulement ont une croissance plus lente. Le taux de croissance originel est réacquis par les souches

ayant acquis des mutations compensatoires tendant à restaurer un superenroulement normal.

Le degré moyen de superenroulement *in vivo* peut donc être important. Toutefois, ce degré n'est pas nécessairement le même partout dans le génome, mais peut fluctuer. Il est possible qu'une augmentation du superenroulement dans des régions particulières puisse favoriser des transitions structurales de l'ADN.

Un exemple d'une telle transition est, par exemple, la séparation des brins facilitée par l'effet du superenroulement négatif sur le sousenroulement de la double hélice. L'excès d'énergie d'une molécule superenroulée négativement peut fournir l'énergie *nécessaire* à la séparation des brins d'ADN.

L'énergie requise pour dérouler l'ADN dépend de la séquence des paires de bases ; elle est, en effet, l'inverse de l'énergie libérée lors de la formation d'une double hélice (cf. figure 3.5). Il faut fournir 12 à 50 kcal/mole pour séparer 10 paires de bases. Le degré réel du superenroulement correspond donc probablement à une quantité suffisante d'énergie pour rendre possible le déroulement d'un petit nombre seulement de paires de bases. Mais de tels effets pourraient être importants, par exemple, pour initier une séparation des brins, lors de la réplication de l'ADN.

À l'extrême le déroulement d'une double hélice droite peut être suivi du réenroulement du double-brin gauche. C'est précisément ce qu'implique la conversion d'une forme d'ADN-B en une forme Z, et, en effet, l'ADN négativement superenroulé prendra plus aisément la forme Z que l'ADN relâché. Dans une molécule d'ADN circulaire contenant

des groupes de répétitions $\begin{smallmatrix} \text{CG} \\ \text{CG} \end{smallmatrix}$, la transition vers la forme Z, ne se produit à des conditions salines isotoniques que lorsque la densité de superenroulement négatif atteint une valeur critique. Quand les résidus C sont méthylés, il faut moins de supertours négatifs pour aider à cette transition.

Plus la zone des doublets $\begin{smallmatrix} \text{CG} \\ \text{CG} \end{smallmatrix}$ sera longue, plus elle aura de facilité à se convertir en forme Z. Ainsi un enchaînement de 32 doublets se convertit à une densité superhélicoïdale de $\sigma = -0,05$, mais il faut une valeur de $-0,07$ pour convertir un enchaînement de 14 doublets.

Il faut une quantité relativement importante d'énergie pour *amorcer* la conversion d'un ADN-B en un ADN-Z ($\Delta G = + 7,7$ kcal/mole), mais une quantité relativement faible ($\Delta G = + 0,45$ kcal/mole pour chaque paire de bases) pour *étendre* la région d'ADN-Z à l'intérieur de la zone

de doublets $\begin{smallmatrix} \text{GC} \\ \text{CG} \end{smallmatrix}$.

L'énergie demandée est du même ordre de grandeur que l'énergie rendue disponible par les superenroulements négatifs ; nous voyons donc que l'apparition du superenroulement est potentiellement une force puissante pour influencer la structure de la double hélice.

Que se passe-t-il à la jonction entre la région d'ADN-Z et celle d'ADN-B ? Apparemment, elle est constituée d'au moins une paire de bases (peut-être plusieurs), qui ne sont plus appariées et ne font plus partie de la double hélice. Cette nécessité de maintenir une séparation des brins, au niveau des jonctions ADN-Z/ADN-B explique en partie la grande quantité d'énergie requise pour produire chaque jonction.

Un autre effet possible du superenroulement est d'engendrer des structures cruciformes dans de l'ADN palindromique. Une quantité considérable d'énergie est nécessaire pour conserver une structure cruciforme, de l'ordre de $\Delta G = + 18$ kcal/mole. Toutefois, il faut une quantité encore plus considérable d'énergie d'activation pour rompre la double hélice, afin de donner naissance à la structure cruciforme, à peu près $\Delta G = + 50$ kcal/mole. De plus, dans des conditions physiologiques, la réaction se produit très lentement. En fait, il semble que le degré de superenroulement qui serait nécessaire pour engendrer une structure cruciforme, soit bien supérieur à celui existant dans une cellule vivante.

POUR EN SAVOIR PLUS

Les structures alternatives de l'ADN ont été passées en revue par **Cantor** (*Cell* 26, 293-295, 1981). La structure de l'ADN-Z a été traitée par **Rich, Nordheim & Wang** (*Ann. Rev. Biochem.* 53, 791-846, 1984). Les conditions pour la formation de structures cruciformes ont été analysées par **Wang** (*Cell* 33, 817-829, 1983).

CHAPITRE 4

ISOLEMENT DU GÈNE

Un chromosome contient une longue tresse ininterrompue d'ADN le long de laquelle se trouvent de nombreux gènes. Les mutations, en créant des modifications dans une séquence de paires de bases, nous fournissent des sites identifiants dans le matériel génétique. Sur la carte de liaison génétique (linkage map), les distances sont représentées par les fréquences de recombinaison. Un chromosome eucaryote peut contenir des milliers de gènes, et au niveau moléculaire, une unité de recombinaison correspond en fait à une très grande région d'ADN.

Pour placer un gène dans un groupe de liaison (linkage), il faut qu'il y ait, à moins de 40 unités de lui, au moins un marqueur précédemment identifié. Pour avoir des repères disponibles dans des intervalles de 40 unités de recombinaison, il faudrait pour un génome humain environ 100 gènes régulièrement espacés. Bien sûr, obtenir 100 repères équitablement espacés, nécessite la cartographie d'un beaucoup plus grand nombre de loci.

Placer les gènes dans un groupe de liaison n'est que le point de départ de leur cartographie. Une localisation plus précise, à l'aide de la génétique formelle, nécessite une plus haute concentration de marqueurs, situation qui n'est rencontrée, chez les eucaryotes supérieurs, que chez le ver nématode (*C. elegans*) et la mouche du vinaigre (*D. melanogaster*). Chez le mammifère dont la carte est la mieux connue, la souris, certaines régions contiennent pratiquement un marqueur toutes les unités de recombinaison, alors que dans d'autres régions, deux marqueurs peuvent être distants de plus de 30 unités.

La cartographie des gènes ne se réduit pas à la seule

analyse génétique. Nous pouvons également utiliser des techniques pour visualiser l'emplacement sur le chromosome des séquences correspondant aux gènes ayant été isolés. Souvent relativement sommaires, ces méthodes nous permettent néanmoins de voir directement où est situé un gène. Elles inversent l'ordre des événements : au lieu d'utiliser la cartographie génétique pour isoler le gène, on utilise le gène isolé pour réaliser la cartographie.

En comparant la longueur totale de la carte génétique de tous les groupes de liaison à la quantité totale d'ADN, nous pouvons avoir une idée de la fréquence moyenne des échanges génétiques par unité de longueur de l'ADN, 50 % de recombinaisons correspondant à la distance moyenne entre des recombinaisons indépendantes. Les valeurs pour plusieurs génomes sont résumées dans le **Tableau 4.1**.

Les bactériophages prennent part à des échanges génétiques *excessivement* fréquents (bien que toute quantification de la fréquence soit seulement *très* approximative). Il est également possible de produire un énorme nombre de descendants, de sorte que, même les recombinaisons les plus rares (des mutations dans des paires de bases contiguës), puissent être mesurées. Ces caractéristiques permettent la cartographie détaillée d'un gène, décrite au Chapitre 1. Le taux d'échange génétique dans la bactérie est également élevé, et il est à nouveau possible d'identifier les recombinants provenant d'un nombre élevé de descendants.

Il est difficile d'appliquer la notion de groupe de liaison aux procaryotes, du fait du taux très élevé de recombinaison.

Tableau 4.1

La fréquence des recombinaisons génétiques est beaucoup moins élevée chez les eucaryotes que chez les procaryotes.

Espèces	Taille du génome haploïde	Unités de carte génétique	Taille de l'unité de recombinaison	Distance moyenne entre les « cross-over »
Phage T4	$1,6 \times 10^5$ pb	800	200 pb	$1,0 \times 10^4$ pb
<i>E. coli</i>	$4,2 \times 10^6$ pb	1750	2 400 pb	$1,2 \times 10^5$ pb
Levure	$2,0 \times 10^7$ pb	4200	5 000 pb	$2,5 \times 10^5$ pb
Nématode	$8,0 \times 10^7$ pb	320	250 000 pb	$1,2 \times 10^7$ pb
Mouche du vinaigre	$1,4 \times 10^8$ pb	280	500 000 pb	$2,5 \times 10^7$ pb
Souris	$3,0 \times 10^9$ pb	1700	1 800 000 pb	$9,0 \times 10^7$ pb
Homme	$3,3 \times 10^9$ pb	3300	1 000 000 pb	$5,0 \times 10^7$ pb

sons, chaque unité correspondant, en fait, à un petit nombre de paires de bases. La carte génétique des procaryotes est dès lors plus ou moins établie au niveau du gène lui-même.

Chez les eucaryotes inférieurs, au moins pour les levures et autres champignons, la fréquence de recombinaison reste élevée ; chez les eucaryotes supérieurs elle est 100 fois plus faible. Localiser des gènes dans des groupes de liaison est réalisable, mais les difficultés de cartographie au niveau moléculaire sont dues à la grande taille de l'unité de recombinaison ($10^5 - 10^6$ pb), et au petit nombre de descendants pour chaque croisement.

A partir des seules données de la cartographie génétique, les perspectives de caractérisation des gènes individuels sont faibles. Supposons qu'un gène de la mouche du vinaigre soit composé de 5 000 paires de bases. Sa longueur correspond à 0,01 % de recombinaison. Pour détecter une recombinaison entre des mutations situées à des extrémités opposées du gène, il serait nécessaire d'isoler une mouche parmi 10 000 descendants. Il en résulte que la cartographie génétique à l'intérieur des gènes d'une mouche du vinaigre est, à la limite, possible. Ce n'est pas réalisable du tout chez la souris, où la fréquence des recombinaisons est plus basse, et le nombre de descendants par croisement, moindre.

Avec quelle exactitude la carte génétique représente-t-elle la longueur réelle du matériel génétique ? Au niveau grossier d'un chromosome, la carte génétique fournit généralement une représentation raisonnable de l'organisation apparente (comme il est montré à la Figure 1.12), bien que la distance de recombinaison entre chaque paire de marqueurs ne reflète pas précisément leur distance physique (voir la Figure 4.10). Bien entendu, l'unité de carte dépend de la fréquence relative de recombinaison ; et puisque cela peut être différent, et de manière frappante, pour chaque espèce, les cartes génétiques d'organismes différents ne peuvent pas être comparées.

Même chez les procaryotes, l'information que l'on peut tirer d'une cartographie poussée de la structure est limitée par la nature même de l'événement de recombinaison. Au niveau de la cartographie interne d'un gène, les fréquences de recombinaison ne sont pas complètement indépendantes du type d'événement utilisé pour cette recombinaison, mais peuvent être influencées assez fortement par la séquence locale d'ADN. En d'autres termes, elles n'obéissent pas aux propriétés idéales d'**indépendance des allèles** décrite au Chapitre 1, mais témoignent d'effets dépendant de la nature des allèles. Notre vision du gène selon la carte génétique se trouve donc être déformée par les caractéristiques des systèmes de recombinaison.

Au niveau moléculaire, les distances sur la carte entre les mutations ne correspondent pas nécessairement aux distances qui les séparent réellement dans l'ADN. L'analyse d'une carte génétique est également limitée par la disponibilité des mutations ; un « blanc » dans la carte pourrait représenter une région dans laquelle on ne trouve pas de mutation, ou être causé par une augmentation locale de la fréquence de recombinaison.

Nous n'avons aucune raison de croire que la situation serait différente chez les eucaryotes si nous étions à même de réaliser une cartographie génétique fine. Donc, pour des raisons à la fois théoriques et pratiques, il nous faut trouver une autre source d'informations. Comment allons-nous établir la structure moléculaire du gène et la rapporter à la structure du produit protéique ? Quelle distance sépare des gènes adjacents et comment allons-nous identifier les régions qui les séparent ?

LES ENZYMES DE RESTRICTION CLIVENT L'ADN EN FRAGMENTS SPÉCIFIQUES

L'objectif final de la cartographie génétique est d'établir la séquence nucléotidique du gène et d'étendre la

séquence aux gènes voisins de chaque côté. Si l'on opère à partir de la protéine, il est possible d'utiliser les techniques d'hybridation des acides nucléiques pour isoler l'ADN correspondant. C'est en fait, maintenant, presque une routine d'isoler un ADN cellulaire si l'on possède le produit protéique correspondant (bien que cela puisse être un long travail si ce produit est rare).

Ayant isolé l'ADN, un pas crucial sur le chemin de l'obtention de sa séquence est de cartographier l'acide nucléique au niveau moléculaire. Une carte physique de n'importe quelle molécule d'ADN peut être obtenue en la fractionnant en des points précis dont les distances respectives peuvent être précisément déterminées. Le fractionnement nécessaire spécifique est rendu possible grâce à certaines **enzymes de restriction** qui reconnaissent des séquences spécifiques et plutôt courtes d'ADN en tant que cibles de clivage.

Chaque enzyme de restriction possède une cible précise dans le double-brin d'ADN, habituellement une séquence spécifique de 4 à 6 paires de bases. L'enzyme coupe l'ADN chaque fois que se présente sa séquence cible. Des enzymes de restriction différentes ont des séquences cibles différentes ; une large variété de ces enzymes (obtenues à partir de bactéries différentes) sont maintenant disponibles. (Elles sont traitées dans le contexte de leur habitat naturel au Chapitre 15.)

Une carte d'ADN obtenue par l'identification de ces points de coupure est appelée une **carte de restriction**. La carte représente une séquence linéaire des sites, où des enzymes de restriction spécifiques trouvent leurs cibles. La distance sur ces cartes se mesure directement en paires de bases (**pb** en abrégé) pour des distances courtes ; les distances plus longues sont données en **kb**, correspondant aux kilopaires de bases dans l'ADN ou aux kilobases dans l'ARN.

D'après la carte de restriction, nous pouvons commencer à déterminer la séquence de l'ADN entre des points suffisamment proches les uns des autres (entre environ 300 pb pour des raisons pratiques). Si l'on choisit correctement ces points, ces régions peuvent être reliées, dans la séquence d'un gène entier et de ses environs. En comparant la séquence d'ADN avec la séquence de la protéine que le gène représente, il est possible de délimiter les régions qui codent pour le polypeptide ; et en étendant la séquence dans les deux sens, de déterminer la distance jusqu'au gène suivant.

En comparant la séquence d'un type d'ADN sauvage avec celle d'un allèle mutant, nous pourrions déterminer la nature de la mutation et l'endroit exact où elle se produit. Cela définit le rapport entre la carte génétique

(entièrement basée sur les sites de mutation) et la carte physique (basée sur, ou même incluant, la séquence d'ADN). Finalement, la carte d'une région d'un génome peut être exprimée en paires de bases d'ADN plutôt que dans les unités relatives de recombinaison de la génétique formelle. En effet, les gènes peuvent à présent être identifiés grâce à la protéine qu'ils produisent, ou même parfois grâce à leur seule séquence et nous ne sommes ainsi plus totalement dépendants des mutations pour nous procurer le matériel brut nécessaire à la construction de la carte d'un génome. Bien entendu, les mutants restent essentiels pour identifier les fonctions des produits d'un gène.

Si nous prenons une molécule d'ADN spécifique et que nous la coupons au moyen d'une enzyme de restriction adéquate, un nombre limité de coupures ont lieu, produi-

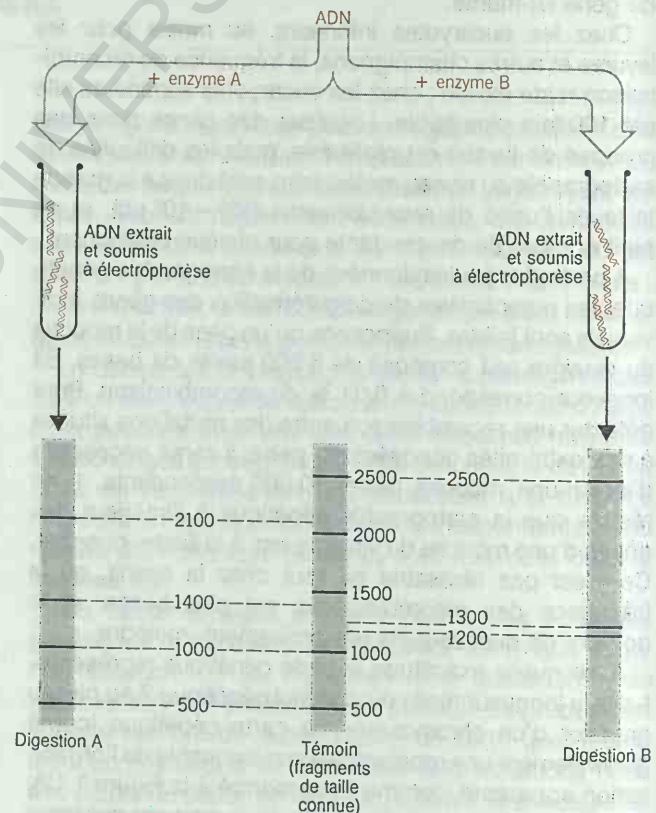


Figure 4.1

L'ADN peut être clivé par des enzymes de restriction en des fragments qui peuvent être séparés par une électrophorèse en gel d'agarose.

Les tailles des fragments distincts engendrés par l'enzyme A (à gauche) ou l'enzyme B (à droite) sont déterminées par comparaison avec les positions des fragments de taille connue, grâce au témoin figurant au centre.

sant des fragments distincts. Ces fragments peuvent être séparés par leur taille au moyen d'une **électrophorèse en gel**. Dans cette technique, l'ADN fragmenté est placé dans un gel d'agarose ou de polyacrylamide. Lorsque l'on fait passer un courant électrique à travers le gel, les fragments se déplacent à une distance qui dépend de leur longueur. Plus le fragment est petit, plus il se déplace vite. (La distance de déplacement dépend du logarithme de la longueur du fragment.)

Ce mouvement produit une série de **bandes**. Chaque bande correspond à un fragment d'une taille spécifique, la taille diminuant vers le bas du gel. La longueur de chaque fragment isolé peut être déterminée en **calibrant** le gel. Cela est réalisé en faisant migrer un témoin en parallèle, dans un autre canal du même gel. Le témoin est composé d'un mélange de fragments standards, tous de taille connue (souvent appelés **marqueurs**). La migration des marqueurs établit le rapport entre la longueur du fragment et la distance parcourue pour un gel donné.

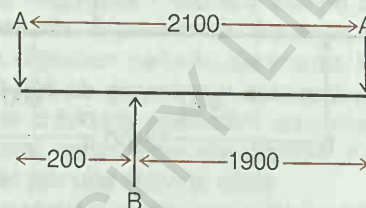
La Figure 4.1 montre un exemple de cette technique. Une molécule d'ADN d'une longueur de 5 000 pb est incubée séparément avec deux enzymes de restriction, A et B. Après la coupure, l'ADN est soumis à électrophorèse, révélant ainsi que l'enzyme A a coupé le substrat d'ADN en quatre fragments (de longueurs respectives 2 100, 1 400, 1 000 et 500 pb) tandis que l'enzyme B a produit trois fragments (dont les longueurs sont 2 500, 1 300 et 1 200 pb). Pouvons-nous progresser d'après ces données pour tracer une carte qui situerait les sites de coupure en des points définis de l'ADN ?

CONSTRUCTION D'UNE CARTE DE RESTRICTION A PARTIR DES FRAGMENTS

Les diagrammes de coupures réalisées par les deux enzymes peuvent être reliés par de nombreux moyens. La Figure 4.2 nous montre le principe de l'analyse par **double digestion**. Dans cette technique l'ADN est fragmenté simultanément par deux enzymes, de même que par chacun d'eux séparément. La façon la plus complète d'utiliser cette technique est d'extraire chaque fragment produit par les digestions individuelles avec, soit l'enzyme A, soit l'enzyme B, et ensuite de le couper avec l'autre enzyme. Les produits de la coupure sont analysés, à nouveau, par électrophorèse.

Chacun des gels de la Figure 4.2 est dénommé selon le fragment qui a été isolé du gel de la Figure 4.1. C'est ainsi que A – 2 100 correspond au fragment de 2 100 pb produit en dégradant la molécule d'ADN d'origine par

l'enzyme A. Lorsque ce fragment est extrait et soumis à l'enzyme B, il est coupé en fragments de 1 900 et 200 pb. Donc l'un des fragments produits par l'enzyme B se situe à 200 pb du plus proche site de coupure de l'enzyme A d'un côté, et de l'autre côté, il se situe à 1 900 pb du site coupé par l'enzyme A. Cette situation est décrite par la carte suivante :



Les flèches verticales indiquent les sites de clivage par l'enzyme A ou l'enzyme B ; les flèches horizontales et les chiffres indiquent les longueurs de chaque fragment.

Un diagramme de coupures coordonnées peut être obtenu par la digestion du fragment B – 2 500 au moyen de l'enzyme A. Il est coupé en fragments de 1 900 et 600 pb. Donc, le fragment de 1 900 pb est produit par des doubles coupures, avec un site A à une extrémité et un site B à l'autre. Il peut être libéré par l'un ou l'autre des fragments, coupés une seule fois (A – 2 100 et B – 2 500), qui le contiennent. Les fragments produits par

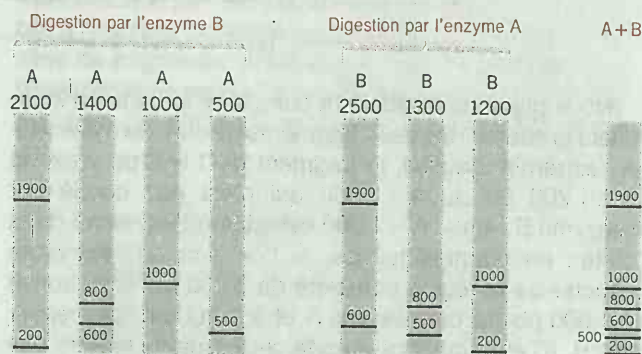
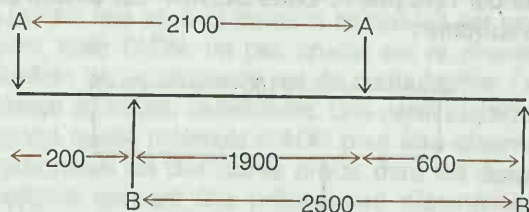


Figure 4.2

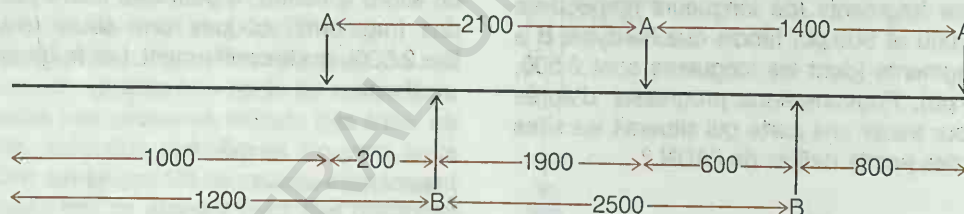
Les doubles digestions définissent les positions des coupures d'une enzyme par rapport à celles de l'autre.

Les quatre fragments obtenus par digestion au moyen de l'enzyme A peuvent être élués du gel montré à la Figure 4.1 ; ensuite ils sont dégradés par digestion au moyen de l'enzyme B (les quatre gels de gauche). De même, les trois bandes initialement produites par l'enzyme B sont isolées et ensuite coupées par l'enzyme A (les trois gels du centre). La double digestion complète est obtenue en soumettant l'ADN entier simultanément aux deux enzymes (gel de droite).

simple coupure doivent ensuite **se chevaucher** dans la région du fragment commun de 1 900 pb qu'ils peuvent engendrer. Pour le montrer, il faut étendre notre carte sur la droite afin d'ajouter un site de coupure de l'enzyme B.



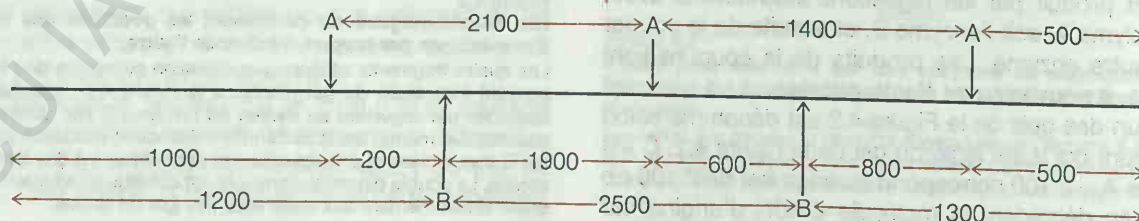
Cette carte illustre un principe important de la cartographie de restriction. Lorsque nous étudions la construction des plus longs fragments à partir des plus petits, nous pouvons nous appuyer sur l'**additivité complète** des longueurs (à l'intérieur des limites expérimentales). Ainsi, le fragment A-2 100 est constitué de deux fragments de 200 pb et 1 900 pb, de même le fragment B - 2 500 est constitué de fragments de 1 900 pb et de 600 pb.



Nous pouvons maintenant compléter la carte en identifiant la source des deux fragments de chaque extrémité. A l'extrémité gauche, le fragment de 1 000 pb vient du B - 1 200 ou du A - 1 000 qui n'est pas coupé par l'enzyme B. Ainsi, A - 1 000 est situé à l'extrémité de la carte ; en d'autres termes, si l'on part de l'extrémité gauche de la région complète de 5 000 pb, il se trouve à 1 000 pb du premier site A et à 1 200 pb du premier site B. (C'est pourquoi la carte ci-dessus ne montre pas de coupure B à gauche, bien qu'au départ nous traitons

Lorsque tous les fragments sont analysés de cette manière, nous pouvons voir que chaque fragment produit en coupant un fragment A d'origine avec l'enzyme B est *également* obtenu par des doubles digestions dans lesquelles un fragment B d'origine a été coupé par l'enzyme A. On peut voir la totalité du profil dans la double digestion complète (le gel à droite de la Figure 4.2), où chaque fragment produit par double digestion est obtenu une seule fois. Grâce à ces données nous pouvons alors localiser les sites de coupure sur une carte sans équivoque.

La clé de la cartographie de restriction réside dans l'utilisation de fragments chevauchants. Grâce au chevauchement de A - 2 100 et de B - 2 500 dans la région centrale de 1 900 pb, nous pouvons relier le site A situé à 200 pb à gauche de la région de 1 900 pb avec le site B situé 600 pb à droite. Il est donc maintenant possible d'étendre la carte plus loin de chaque côté. Le fragment de 200 pb à gauche est également produit en coupant B - 1 200 par l'enzyme A, donc le prochain site B doit se situer à 1 000 pb sur la gauche. Le fragment 600 pb à droite est également obtenu en coupant A - 1 400 par l'enzyme B, le prochain site A devant donc se trouver à 800 pb vers la droite. Cela nous donne la carte suivante :



dans notre analyse cette extrémité comme un site de coupure B.)

A l'extrémité droite de la carte, le fragment de 800 pb est obtenu par double digestion en coupant B - 1 300 par l'enzyme A, nous devons donc ajouter un fragment de 500 pb vers la droite. C'est le fragment terminal, comme nous le voyons par sa présence en tant que A - 500 dans la simple digestion de A. Ainsi, notre carte complète prend la forme suivante :

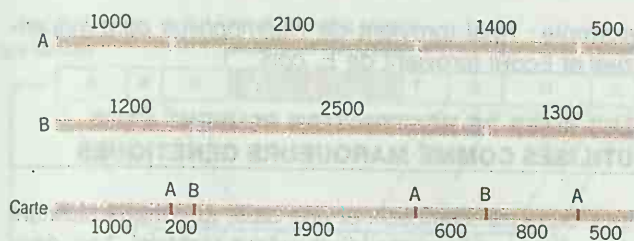


Figure 4.3

Une carte de restriction est une séquence de sites séparés par des distances précises sur l'ADN.

La carte permet de repérer les sites coupés par les enzymes A et B comme le montrent les fragments isolés produits par simple ou double digestion.

Nous avons maintenant réalisé une carte de restriction qui couvre entièrement la région de 5 000 pb. Cela est récapitulé sous la forme plus courante en **Figure 4.3**. La carte montre les positions auxquelles les enzymes de restriction spécifiques coupent l'ADN. Les distances entre les sites de coupure sont mesurées en paires de bases. Ainsi l'ADN est divisé en une série de régions de longueurs données qui s'étendent entre les sites reconnus par les enzymes de restriction.

La construction réelle d'une carte de restriction implique habituellement le recours à plusieurs enzymes ; il devient donc nécessaire de débrouiller le profil relativement complexe des fragments chevauchants, produits par les différentes enzymes. Plusieurs techniques supplémentaires sont utilisées pour aider à la cartographie.

Une des approches consiste à réaliser des **digestions partielles** en utilisant des conditions où une enzyme ne reconnaît pas chaque cible, dans chaque molécule d'ADN, mais, au lieu de cela, clive plus ou moins les différents sites. Il faudra donc fixer les conditions pour

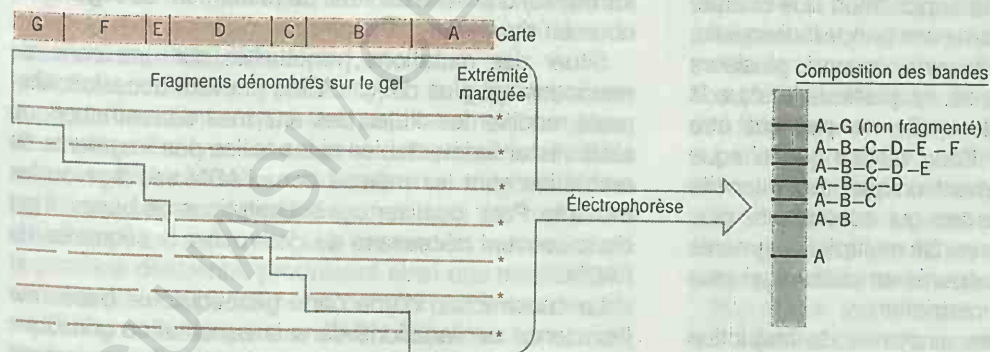


Figure 4.4

Lorsque des fragments de restriction sont détectés grâce à une extrémité marquée, chacun indique directement la distance entre un site de coupure et l'extrémité. L'augmentation de la longueur des fragments successifs dépend de la distance entre les sites de restriction adjacents.

que l'enzyme ait, par exemple, une probabilité de 50 % de couper chacun de ses sites de reconnaissance.

Avec cette technique, l'enzyme A pourrait opérer un clivage partiel des fragments de 3 500, 3 100 et 1 400 pb à partir du segment d'ADN cartographié à la Figure 4.3. En comparant les produits des digestions partielles au schéma des digestions complètes, il est possible de localiser les fragments 1 000 et 2 100 pb comme étant des composants adjacents du fragment de 3 100 pb obtenu par digestion partielle. De même, les fragments 2 100 et 1 400 pb pourraient, contigus, constituer le fragment de 3 500 pb également issu d'une digestion partielle. Ainsi, cette technique permet d'ordonner les sites coupés par une enzyme seule.

Une autre technique utile est le **marquage des extrémités** par laquelle les extrémités d'une molécule d'ADN sont marquées par un phosphate radioactif (certaines enzymes peuvent ajouter des résidus phosphate, spécifiquement aux extrémités 5' ou 3'). Cela permet aux fragments qui contiennent les extrémités d'être directement identifiés par leur marquage radioactif. Ainsi, dans le fragment de la préparation A, A - 1 000 et A - 500 seraient immédiatement placés aux extrémités opposées de la carte, de même que les fragments B - 1 200 et B - 1 300.

En combinant la digestion partielle avec le marquage des extrémités, une série de sites coupés par une enzyme peut être directement localisée en fonction de l'extrémité. La **Figure 4.4** montre que si les fragments ne sont identifiés que par leurs extrémités marquées, ceux coupés à l'intérieur de la molécule sont ignorés, tandis qu'une série de fragments, produits par coupure partielle, identifie la distance entre chaque site de coupure et l'extrémité marquée. Si l'extrémité gauche du fragment de 5 000 pb

de la Figure 4.3 était marquée, une digestion partielle avec l'enzyme A permettrait immédiatement de localiser les sites de coupure à 1 000, 3 100 et 4 500 pb de l'extrémité.

Pour analyser les résultats obtenus par coupure par des enzymes multiples, les fragments sont souvent comparés au moyen de la technique d'hybridation de l'acide nucléique. Cela renforce considérablement l'analyse. Par exemple, en réalisant la carte de la Figure 4.3, nous déduisons que les fragments A - 2 100 et B - 2 500 doivent se chevaucher puisqu'ils produisent tous les deux un fragment de 1 900 pb par double digestion. Cependant, lors de l'analyse d'un ADN plus long, il pourrait exister plusieurs fragments de tailles identiques. L'hybridation nous permet de voir directement si deux fragments se chevauchent, ainsi, A - 2 100 et B - 2 500 devraient s'hybrider.

Toutes ces techniques exigent que nous ayons à notre disposition un jeu complet de fragments uniques de restriction, qui à eux tous représentent la région entière d'ADN à cartographier. Dans certaines circonstances, ceci n'est pas possible.

Si plusieurs sites, pour une seule enzyme, sont *très proches les uns des autres* (disons dans les 50 pb), cela peut produire de très petits fragments qui seront perdus dans le gel d'agarose. Bien sûr, l'addition des poids moléculaires en serait faussée, mais cela ne serait pas nécessairement détecté, puisqu'il existe toujours de modestes erreurs expérimentales dans l'appréciation des tailles des fragments. Cette difficulté est surmontée en utilisant plusieurs enzymes de restriction dans la réalisation d'une carte, ainsi il y a suffisamment de chevauchements pour que nous soyons sûrs que tout l'ADN est couvert.

Une complication vient de la supposition que chaque fragment est vu, sur le gel, comme une bande individuelle. Ceci n'est, évidemment, pas vrai, quand plusieurs fragments ont la même taille, ou, en pratique lorsque la taille des fragments est telle qu'ils ne peuvent être séparés par électrophorèse. Pour vérifier que chaque bande correspond à un fragment unique, son intensité peut être quantifiée. Les bandes qui contiennent plus d'un fragment sont plus intenses. De multiples fragments de même taille peuvent être séparés en utilisant un plus grand nombre d'enzymes de restriction.

Un point final à propos des enzymes de restriction concerne leur nomenclature. Les enzymes sont désignées par une abréviation de trois lettres (parfois quatre) qui identifie leur origine. On ajoute des chiffres romains pour distinguer plusieurs enzymes de même origine. Par

exemple : HpaI (provient de *Haemophilus para-influenzae*) et EcoRI (provient de *E. coli*).

LES SITES DE RESTRICTION PEUVENT ÊTRE UTILISÉS COMME MARQUEURS GÉNÉTIQUES

Une carte de restriction et une carte génétique donnent des visions différentes d'un même matériel. Les deux cartes peuvent être rapprochées, pour éclairer la base physique sous-jacente à la carte génétique.

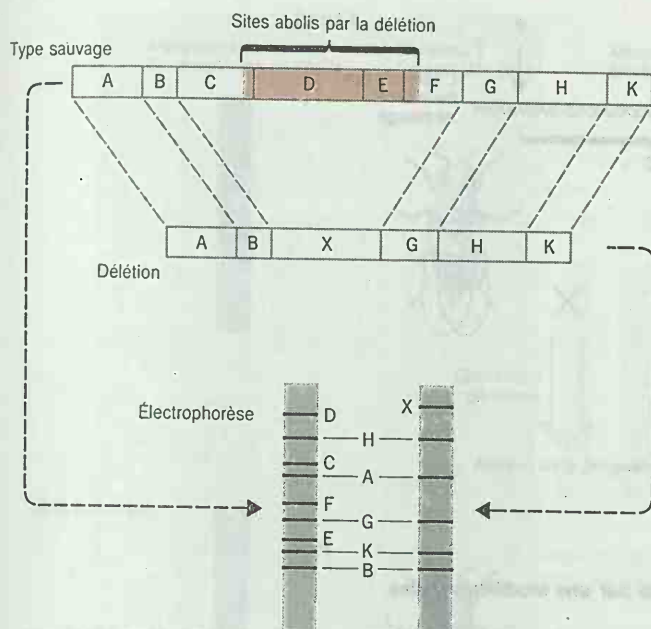
Une carte de restriction identifie une série linéaire de sites dans l'ADN, séparés les uns des autres par une distance réelle le long de l'acide nucléique. *Il est possible d'obtenir une carte de restriction pour chaque séquence d'ADN, sans que l'on ait identifié ses mutations, ou en fait, sans que l'on connaisse, du tout, sa fonction.*

Une carte génétique identifie une série de sites où se produisent les mutations. *L'existence des sites dépend du fait que des changements dans la séquence des bases ont altéré le phénotype. La distance entre les sites est déterminée par les fréquences de recombinaison, qui sont elles-mêmes influencées par les conditions locales.*

Pour opérer un rapprochement entre la carte de restriction et la carte génétique, il nous faut comparer les cartes de restriction du type sauvage avec celles des ADN mutants correspondants. Le rapport entre carte génétique et carte physique peut être déterminé sur la seule base de la carte de restriction lorsque les mutants produisent d'importants changements dans la carte génétique. Par exemple, une petite délétion ou insertion affectant la carte génétique provoquerait une diminution ou une augmentation correspondante de la taille du fragment de restriction dans lequel elle se produit. Une délétion (ou une insertion) plus longue peut déplacer (ou introduire) une série de sites de restriction. La **Figure 4.5** nous en montre un exemple.

Situer des mutations ponctuelles sur une carte de restriction est plus difficile. Elles peuvent occasionnellement modifier les cibles des enzymes de restriction, ou alors rester indétectables si les tailles des fragments de restriction sont les mêmes dans l'ADN sauvage ou les mutants. Pour localiser ces substitutions de bases, il est donc souvent nécessaire de déterminer la séquence de l'ADN.

La construction d'une carte génétique est basée sur l'existence de variations dans la constitution génétique de la population. La coexistence de plus d'un variant est appelée **polymorphisme génétique**. Chaque site pour lequel on peut définir des allèles multiples en tant que composants stables de la population est, par définition,

**Figure 4.5**

Une délétion fait disparaître une série de sites de restriction et fait fusionner les fragments de restriction contenant ses extrémités.

La délétion élimine entièrement les fragments D et E qui faisaient partie des fragments de restriction du type sauvage. Les fragments C et F du type sauvage sont également absents du fragment délété ; on obtient alors un nouveau fragment contenant une partie de leur séquence. Les fragments qui représentent les régions de part et d'autre de la délétion, A et B sur la gauche, G, H et K sur la droite, restent exactement les mêmes dans les produits de digestion du fragment sauvage ou du fragment modifié.

Une insertion aurait exactement l'effet inverse ; un mutant contiendrait des fragments supplémentaires qui ne seraient pas présents dans le type sauvage.

polymorphe. Par exemple, la mouche du vinaigre *D. melanogaster* est polymorphe pour la série d'allèles du locus *blanc* (*white*), W^+ , W^i , W^a , etc.

Si l'on considère le phénotype, le polymorphisme implique l'existence dans une population de mouches, d'un allèle de type sauvage et d'une série d'allèles mutants. Réfléchissons aux bases d'un polymorphisme parmi des allèles mutants. Ils possèdent des mutations qui altèrent le produit du gène en rendant la fonction de la protéine déficiente, produisant ainsi une modification dans le phénotype. Si nous comparons les cartes de restriction ou les séquences d'ADN des allèles concernés, elles seront également polymorphes en ce sens que chaque carte ou séquence sera différente des autres.

Bien que cela ne soit pas évident d'après le phénotype, le type sauvage peut lui-même être polymorphe. Il peut

exister de multiples versions d'un allèle de type sauvage qui se distinguent entre elles par des différences de séquence n'affectant pas leur fonction et qui ne sont donc pas détectées sous forme de variantes phénotypiques. Si l'on considère son génotype, une population de mouches peut avoir un polymorphisme abondant ; de nombreuses séquences variantes peuvent exister dans le locus *w* ; certaines étant manifestes parce qu'elles affectent le phénotype, d'autres cachées parce qu'elles n'ont pas d'effet visible.

Certains polymorphismes dans le génome peuvent être détectés en comparant les cartes de restriction d'individus différents. Le critère est un changement dans la répartition des fragments produits par clivage par une enzyme de restriction. La **Figure 4.6** montre que lorsqu'un site est présent dans le génome d'un individu et absent chez un autre, un clivage supplémentaire dans le premier génome produira deux fragments correspondant au fragment unique du second génome.

Puisque la carte de restriction est indépendante de la fonction du gène, un polymorphisme à ce niveau peut être détecté *indépendamment du fait que le changement de séquence affecte ou non le phénotype*. En fait, une minorité seulement des polymorphismes de sites de restriction dans un génome affectent réellement le phénotype. La majorité peut impliquer des changements de séquence qui n'ont pas d'effet sur la production des protéines (par exemple parce qu'ils sont situés entre les gènes).

Une différence entre les cartes de restriction de deux individus peut être utilisée comme marqueur génétique, exactement de la même manière qu'un autre marqueur. Au lieu d'examiner certaines caractéristiques du phénotype, nous décidons directement que le génotype sera révélé par la carte de restriction. La **Figure 4.7** montre le pedigree d'un polymorphisme de restriction qui se poursuit sur trois générations. Il suit le principe de la ségrégation Mendélienne (déjà illustrée à la **Figure 1.3**) au niveau des fragments marqueurs d'ADN.

La fréquence de recombinaison peut être mesurée entre un marqueur de restriction et un marqueur visible phénotypiquement comme nous le montre la **Figure 4.8**. Ainsi une carte génétique peut contenir des marqueurs génotypiques et des marqueurs phénotypiques.

Parce que les marqueurs de restriction ne sont pas que le reflet des changements du génome affectant le phénotype, ils nous fournissent la base d'une technique extrêmement efficace pour identifier les loci génétiques au niveau moléculaire. Un problème courant concerne une mutation dont les effets sont connus sur le phénotype-

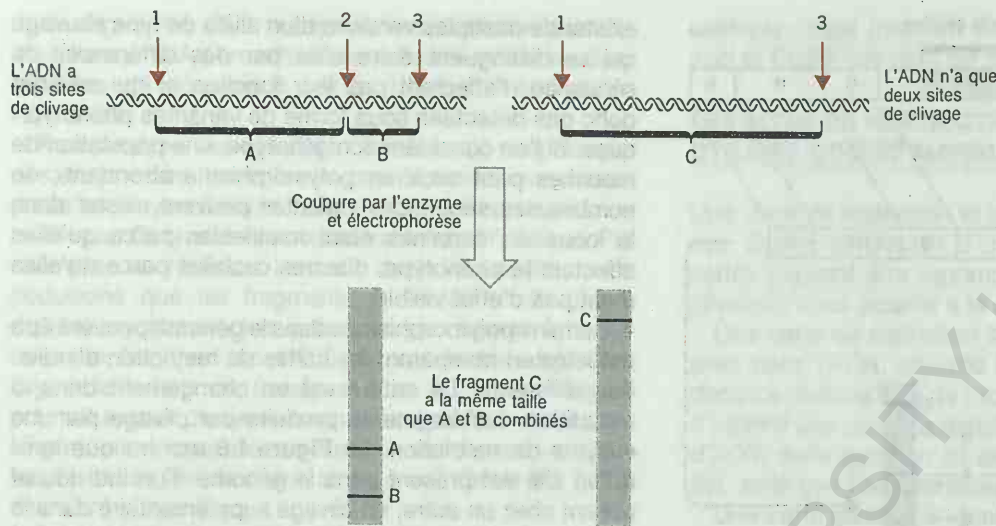


Figure 4.6

Un changement dans l'ADN, affectant un site de restriction, est détecté par une modification des fragments de restriction.

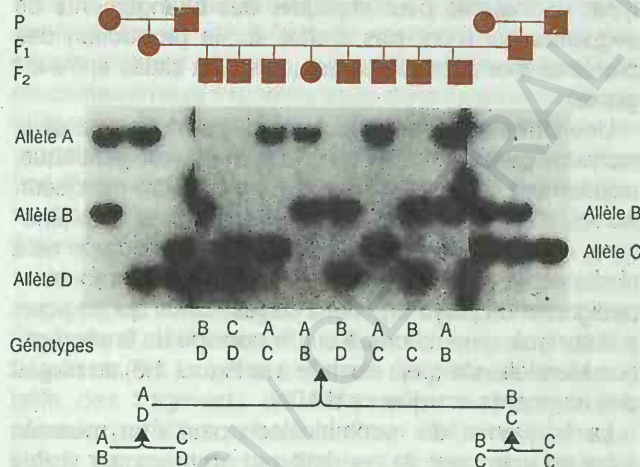


Figure 4.7

Les polymorphismes de restriction sont hérités, selon les lois Mendéliennes. Quatre allèles pour un marqueur de restriction sont présents dans toutes les combinaisons possibles et ségrègent indépendamment à chaque génération.

A la génération P, les trois premiers individus sont hétérozygotes (AB ; CD ; BC) et le dernier est homozygote (CC). Leurs enfants appartenant à la génération F1, sont AD et BC. Ainsi chaque individu de la génération F2 acquiert soit un A soit un D d'un des deux parents, et soit un B soit un C de l'autre parent.

pe, où le locus génétique concerné peut être situé sur une carte génétique, mais pour lequel nous n'avons aucune connaissance de la protéine ou du gène correspondant. Certaines maladies humaines importantes tombent dans cette catégorie. Par exemple, la fibrose kystique, la chorée de Huntington et de nombreuses autres maladies graves ou fatales possèdent une hérédité Mendélienne, mais la nature moléculaire de la fonction mutante est inconnue, et le restera probablement jusqu'à ce que nous soyons capables de caractériser le gène responsable.

Si les polymorphismes de restriction se produisent au hasard dans le génome, certains pourraient intervenir près du gène atteint. Nous pouvons identifier de tels marqueurs de restriction grâce à leurs étroite liaison avec le phénotype mutant. Si nous comparons les cartes de restriction d'ADN de patients souffrant d'une maladie, avec l'ADN de gens normaux, nous pouvons trouver qu'un site de restriction particulier est toujours présent (ou toujours absent) chez les patients.

La Figure 4.9 en montre un exemple hypothétique. Cette situation correspond à une liaison (linkage) de 100 % entre le marqueur de restriction et le phénotype. Cela implique que le marqueur de restriction serait proche du gène mutant, qu'il n'en serait jamais séparé par recombinaison.

L'identification d'un tel marqueur a deux conséquences importantes. Elle pourrait nous conduire à l'isolement du gène. Elle pourrait également nous offrir un moyen de diagnostic permettant de détecter la maladie.

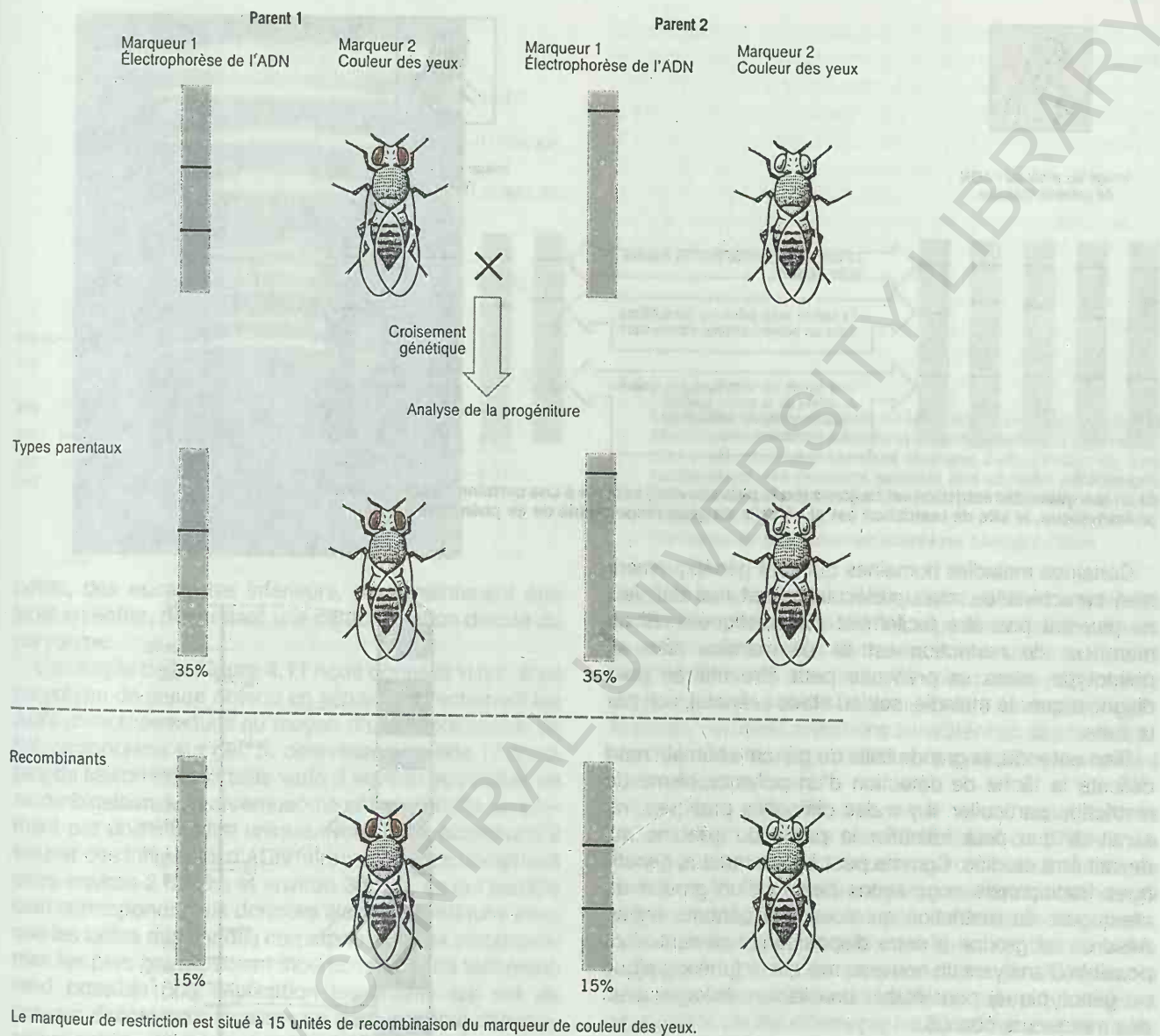


Figure 4.8

On peut utiliser un polymorphisme de restriction comme marqueur génétique pour mesurer la distance de recombinaison avec un marqueur phénotypique (tel que la couleur des yeux).

Notons que la figure simplifie la situation réelle en ne montrant que les bandes d'ADN correspondant à un génome haploïde ; en réalité, il devrait y avoir également des bandes correspondant à l'allèle de l'autre génome de cet organisme diploïde.

Si les deux loci ne se recombinent que rarement ou même jamais, le marqueur doit se trouver relativement près du gène sur la carte génétique. Bien que « relativement proches » en termes génétiques puisse signifier une

distance substantielle en paires de bases d'ADN (voir Tableau 4.1), cela nous fournit néanmoins un point de départ à partir duquel nous serons peut-être capables de progresser le long de l'ADN jusqu'au gène lui-même.

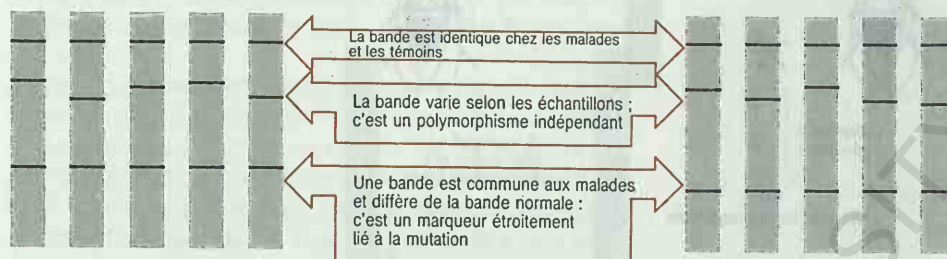
Image du profil de l'ADN
de patients maladesImage des témoins
normaux

Figure 4.9

Si un marqueur de restriction est toujours (ou le plus souvent) associé à une certaine caractéristique phénotypique, le site de restriction est situé près du gène responsable de ce phénotype.

Certaines maladies humaines qui sont génétiquement bien caractérisées, mais moléculairement mal définies, ne peuvent pas être facilement diagnostiquées. Si un marqueur de restriction est lié de manière sûre au phénotype, alors sa présence peut être utilisée pour diagnostiquer la maladie, soit au stade prénatal, soit par la suite.

Bien entendu, la grande taille du génome humain rend délicate la tâche de détection d'un polymorphisme de restriction particulier. Il y a des difficultés pratiques, ne serait-ce que pour identifier la partie du génome qui devrait être étudiée. Comme pour les marqueurs génétiques traditionnels nous avons besoin d'un groupe de marqueurs de restriction qui couvre le génome entier. Avec un tel groupe à notre disposition, il devient alors possible d'analyser un nouveau marqueur (phénotypique ou génotypique) pour établir une liaison (linkage) avec des marqueurs connus.

Grâce à cette technologie nous réalisons un progrès notable vers l'identification des gènes de nombreuses maladies humaines. Par exemple, la **Figure 4.10** montre une carte de liaison du chromosome X humain, basée sur environ 20 polymorphismes de restriction, qui permet à chaque locus d'une maladie liée à l'X d'être cartographié à moins de 10 unités de recombinaison (environ 10 000 kb) d'au moins deux marqueurs.

Nous aimerions rapporter réellement la carte génétique à la séquence chromosomique de l'ADN mais il est difficile d'isoler de longs segments d'ADN. Même le plus petit chromosome humain contient environ 50 000 kb d'ADN. Cependant l'ADN des chromosomes relativement

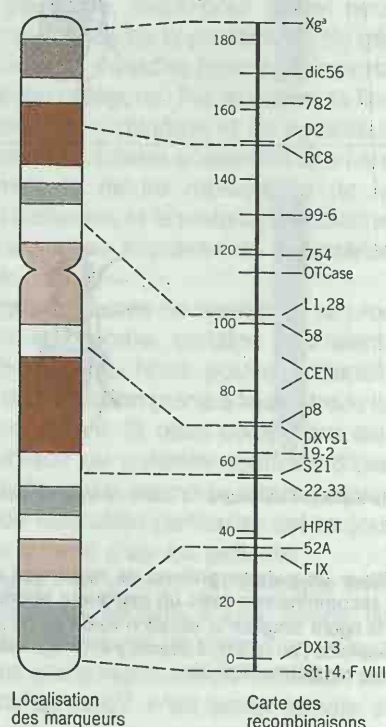


Figure 4.10

Une carte de liaison du chromosome X humain a été réalisée d'après les polymorphismes de restriction.

Notons que les distances entre les marqueurs dont les positions physiques sont connues (montrés à gauche) ne correspondent pas exactement aux distances mesurées par recombinaisons.

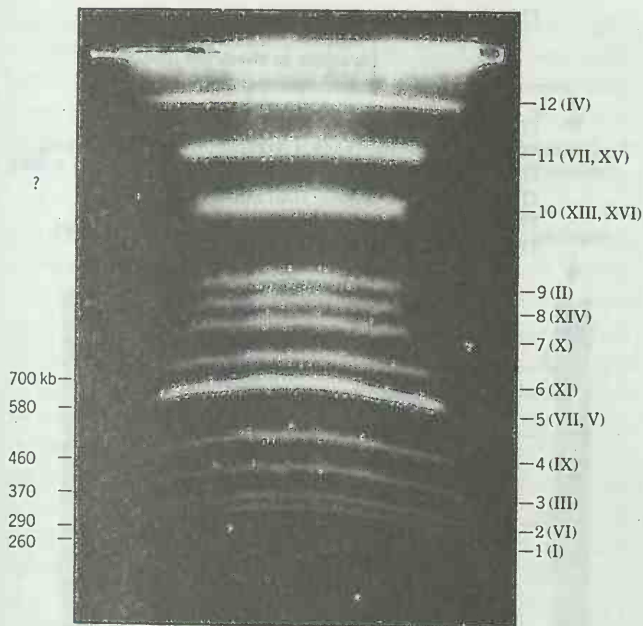


Figure 4.11

Les ADN chromosomiques de la levure *S. cerevisiae* peuvent être séparés en 12 bandes dont 9 représentent des chromosomes isolés et 3 représentent chacune 2 chromosomes. Les composants des doublets peuvent être séparés par d'autres gels. Deux chromosomes ne sont pas identifiés par cette technique.

Photographie gracieusement fournie par Maynard Olson.

petits, des eucaryotes inférieurs, peut maintenant être isolé en entier, permettant une caractérisation directe du caryotype.

L'exemple de la **Figure 4.11** nous donne la vision d'un caryotype de levure obtenu en séparant directement les ADN chromosomiques au moyen d'une modification de l'électrophorèse sur gel. *S. cerevisiae* possède 17 groupes de liaison dont la taille varie d'environ 400 unités de recombinaison à un chromosome entier identifié simplement par une mutation unique. Nous nous attendions à trouver des longueurs d'ADN chromosomique comprises entre environ 2 000 kb et environ 300 kb, ce qui semble bien correspondre aux données électrophorétiques (bien que les tailles des bandes correspondant aux chromosomes les plus grands soient incertaines). Cette technique rend possible une application importante qui est de pouvoir directement localiser un gène sur tout chromosome en identifiant la bande d'ADN avec laquelle il s'hybride. En utilisant des souches de levure possédant des translocations génétiques connues, la cartographie à l'intérieur des régions chromosomiques devient possible. Avec cette technique nous voyons pour la première fois, la possibilité de rapporter directement une carte de liaison à l'ADN chromosomique.

OBTENIR LA SÉQUENCE D'ADN

Par la spécificité de l'appariement des bases, la double hélice maintient une structure constante indépendam-

ment de sa séquence spécifique. La différence entre les molécules individuelles d'ADN, repose sur les séquences spécifiques des paires de bases, et non sur de profondes modifications de structure. De très petites différences dans une séquence d'ADN peuvent être hautement significatives, certaines mutations consistent en des modifications dans la séquence d'ADN qui peuvent être seulement d'une paire de bases.

Comment peut-on déterminer avec exactitude la séquence d'une paire de nucléotides de l'ADN ? La première approche du séquençage d'un acide nucléique suivait la technique du séquençage des protéines ; il s'agissait de fractionner la molécule en petits fragments, de déterminer la composition en bases, et d'obtenir des fragments chevauchants permettant de déduire la séquence exacte. Avec les protéines cela est possible à cause de la variété offerte par les 20 acides aminés ; avec les acides nucléiques cela pose un problème parce qu'il n'y a que quatre bases. Les techniques anciennes étaient donc réduites à ne déterminer les séquences que de très courts acides nucléiques.

Mais à présent il est possible de déterminer les séquences d'ADN *directement*. En fait, les séquences d'ADN peuvent être obtenues beaucoup plus rapidement que les séquences de protéines.

Tous les protocoles de séquençage partagent un principe général. Chacun vise à produire une série de molécules d'ADN simple-brin chacune plus longue d'une base que la précédente. Les molécules d'ADN de *même séquence* mais dont la longueur diffère ne serait-ce que

d'une base à l'extrémité, peuvent être séparées par électrophorèse sur gel d'acrylamide. Des bandes correspondant aux molécules de longueurs croissantes forment une échelle qui peut être lue sur environ 300 nucléotides.

Deux types d'approches ont été utilisés pour obtenir ces séries de bandes. La première, consiste à utiliser les réactions chimiques qui fragmentent l'ADN en des bases individuelles. La seconde, est de synthétiser l'ADN *in vitro* de sorte que la réaction se termine spécifiquement à la position correspondant à une base donnée.

Pour déterminer la séquence de la molécule d'après l'une ou l'autre approche, on la soumet au protocole approprié dans quatre réactions séparées, chaque réaction étant spécifique d'une des bases. Les produits sont soumis à électrophorèse dans quatre canaux parallèles du même gel. La bande correspondant à une longueur spécifique n'apparaît que dans un seul des quatre canaux. Le canal permet d'identifier le nucléotide présent à la position correspondante dans l'ADN. On progresse d'une bande à la bande de taille suivante, ainsi la série de nucléotides peut être « lue » selon les canaux dans lesquels apparaissent les bandes. Cela nous donne la séquence d'ADN.

Dans la méthode chimique un fragment d'ADN, préalablement marqué à une extrémité (unique) avec du ^{32}P radioactif, est coupé spécifiquement à l'une des quatre bases. La clé de l'utilisation du clivage pour le séquençage est de s'assurer qu'il n'y a qu'une coupure partielle à chaque site sensible, apparaissant avec une probabilité de, disons, 1 à 2 %.

Ainsi, commençons-nous donc avec une préparation qui contient un grand nombre de molécules identiques, toutes marquées à la même extrémité. Lorsque la préparation est soumise au traitement chimique, chaque molécule est susceptible d'être coupée en un petit nombre seulement de sites où est présente la base correspondante. Mais dans la préparation globale de molécules d'ADN, chaque position sera attaquée à un moment ou à un autre comme il est montré à la Figure 4.12.

Les molécules traitées sont soumises à électrophorèse, et celles qui portent le marquage radioactif sont repérées (par autoradiographie). Chacune formera une bande dont la longueur dépend de la distance entre l'extrémité marquée et le site de coupure. On obtiendra donc une série complète de bandes, à chaque traitement chimique, marquant chaque position où se trouve la base spécifique par rapport à l'extrémité de la molécule.

Dans la méthode d'origine (Maxam-Gilbert), la spécificité pour les purines était fournie par l'utilisation de diméthylsulfate. Ce réactif méthyle la position N7 de la

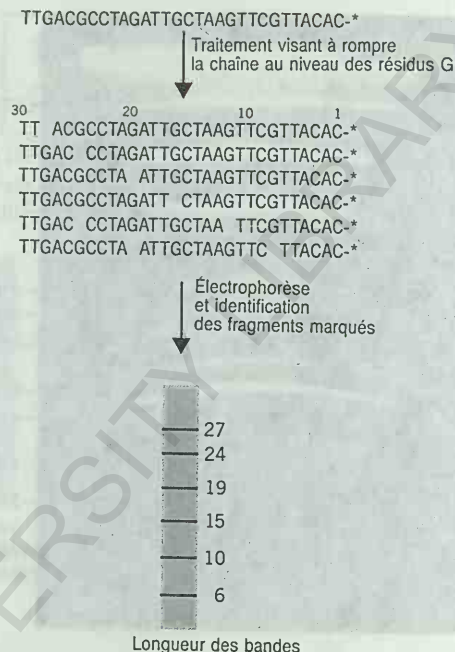


Figure 4.12

Les positions d'une base donnée dans un ADN simple-brin peuvent être déterminées par l'électrophorèse des fragments dont les extrémités sont marquées après les avoir traités par le réactif approprié.

Dans cet exemple, la chaîne est rompue au niveau des résidus G. Certaines chaînes ne sont brisées qu'une fois, d'autres plus d'une fois, mais dans chaque cas, le fragment qui est révélé est celui dont l'extrémité est marquée (indiqué par l'astérisque). Toutes ensemble, les molécules marquées se répartissent de leur extrémité commune à chaque position où un G était présent dans l'ADN d'origine. Après électrophorèse, chaque molécule marquée donne une bande dont la position sur le gel indique la longueur.

guanine environ 5 fois plus efficacement que la position N3 de l'adénine. S'il est chauffé, l'ADN perd la base méthylée, et cela engendre une rupture dans la chaîne polynucléotidique. La réaction étant plus efficace avec la guanine, les résidus G produisent des bandes foncées et les résidus A des bandes claires (c'est-à-dire que le clivage de A a été moins fréquent). Le schéma peut être inversé si, au lieu de chauffer, on utilise de l'acide pour libérer les bases méthylées. Ainsi G et A peuvent être identifiés séparément par des réactions spécifiques.

Les pyrimidines sont analysées par un traitement à l'hydrazine, qui agit avec la même efficacité sur la cytosine et la thymine ; mais il est possible de les différencier en milieu fortement salin où seule la cytosine réagit. Deux séries de bandes sont donc produites, une représentant exclusivement C, et une représentant la combinaison de C + T.

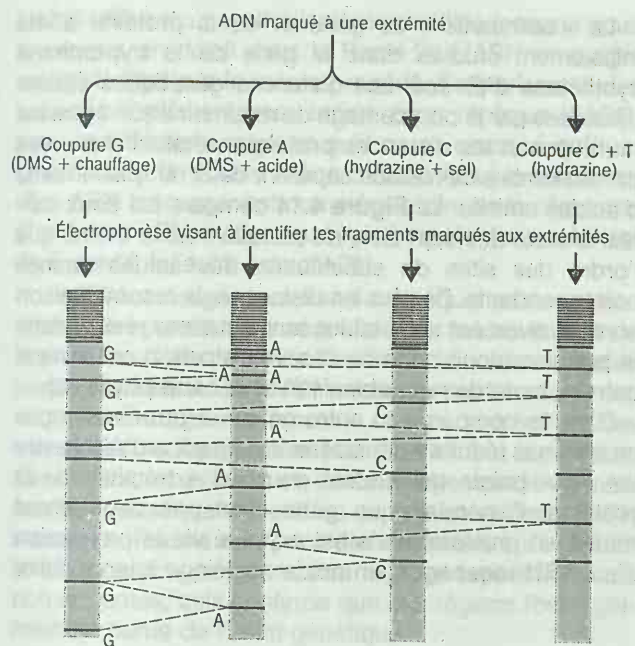


Figure 4.13

L'ADN peut être séquencé par un jeu de réactions qui ensemble permettent de spécifier chacune des quatre bases.

On effectue parallèlement quatre réactions, et les produits sont soumis à électrophorèse. La séquence est lue en partant du bas du gel et en suivant successivement les bandes, comme nous l'indique le zig-zag. G, A et C sont identifiés par les trois gels de gauche, C + T apparaissent eux sur la droite, la différence entre les bandes C et C + T révèle donc les positions du T. La séquence d'ADN est

extrémité-N₂₀GAGCATGACGGTAGCTAGAGTA.....

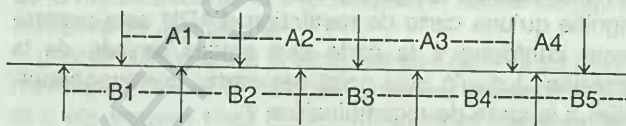
En bas du gel les plus petits fragments ont habituellement une longueur d'environ 20 bases, qui est la distance la plus proche de l'extrémité marquée que nous puissions atteindre. Les bandes deviennent progressivement plus proches en remontant le gel, jusqu'à ce que, en haut du gel, la séquence soit illisible.

Un jeu de quatre mélanges réactionnels est préparé comme il est indiqué à la Figure 4.13. Les produits des réactions sont analysés en parallèle, par électrophorèse. Chacune des quatre réactions contient une série de bandes qui localisent les molécules dans lesquelles la coupure s'est produite aux divers sites où se trouvait la base concernée. La séquence d'ADN est obtenue en progressant d'une bande à la bande suivante plus longue d'une base. Les plus petites bandes sont situées en bas du gel, la séquence d'ADN est donc lue de bas en haut.

Quelle est l'exactitude d'une séquence d'ADN ? Le jeu de réactions nous donne la séquence d'un brin d'ADN. Le risque d'une erreur occasionnelle existe toujours.

Cependant il est possible d'améliorer l'exactitude en séquençant indépendamment les deux brins d'ADN. Chaque fois qu'il existe des endroits non complémentaires, on peut considérer qu'il y a une erreur qui peut être corrigée. Avec cette précaution les erreurs dans la séquence devraient être très rares.

Les molécules d'ADN d'environ 300 nucléotides peuvent être séparées sur des gels d'acrylamide. Comment passe-t-on de la détermination de la séquence de ces fragments, relativement courts, à celle d'une région plus longue ? Exactement comme avec la carte de restriction, la clé en est l'utilisation des fragments chevauchants. Considérons une série de fragments adjacents, ordonnés sur la carte de restriction suivante :



Supposons que nous déterminions la séquence de chacun des fragments de la série A. Nous savons d'après la carte de restriction qu'ils sont adjacents, leurs séquences individuelles devraient donc en théorie s'additionner directement et nous donner la séquence entière. Mais pour être certain qu'il n'y a pas de petits fragments manquants, provoqués par des sites de coupure très proches d'une même enzyme (supposons qu'il y ait des sites de coupure séparés de 10 pb entre A2 et A3), il est essentiel de séquencer des fragments couvrant les points de jonction. Ainsi, avons-nous donc besoin d'une autre série de fragments, dans laquelle tous les points de jonction entre les fragments A sont intacts. Ce sera possible grâce à la série de fragments B, dont chaque fragment contient la séquence de chaque côté du point de jonction entre deux fragments A.

De la même manière que nous réalisons une carte de restriction à partir des fragments chevauchants nous pouvons construire la séquence globale d'un ADN en utilisant les extrémités chevauchantes de chacun des fragments séquencés individuellement. Supposons, par exemple, que les extrémités des fragments A1 et A2 possèdent les séquences suivantes :

A1..... CGTAGGGTCAAGTCAT

AATGCTGCCCAA...A2

Dans le fragment B2 nous trouverons la séquence :

CGTAGGGTCAAGTCATAATGCTGCCCAA

dans laquelle la séquence reproduit en un fragment continu les extrémités des fragments A1 et A2.

LES GÈNES PROCARYOTES ET LES PROTÉINES SONT COLINÉAIRES

En comparant la séquence nucléotidique d'un gène avec la séquence des acides aminés d'une protéine, on peut déterminer directement si le gène et la protéine sont **colinéaires** : c'est-à-dire, si la séquence nucléotidique du gène correspond exactement à la séquence des acides aminés de la protéine. Dans les bactéries et leurs virus, il y a une homologie totale. Chaque gène contient une séquence continue d'ADN dont la longueur correspond directement à la taille de la protéine qu'il représente. Il faudra une séquence de 3N paires de bases pour coder une protéine de N pour acides aminés.

L'équivalence du gène procaryote et de son produit, signifie qu'une carte de restriction d'ADN sera exactement conforme à la carte des acides aminés de la protéine. Jusqu'à quel point ces cartes correspondent-elles à la carte de recombinaison ?

La « colinéarité » du gène et de la protéine a été initialement étudiée dans le gène de la tryptophane synthétase d'*E. coli*. Les distances génétiques étaient mesurées par le pourcentage de recombinaison entre les mutations et les distances protéiques étaient mesurées par le nombre de résidus séparant deux remplacements d'acides aminés. La **Figure 4.14** compare les deux cartes. L'ordre des sept sites de mutation est le même que l'ordre des sites de substitution des acides aminés correspondants. De plus les distances de recombinaison sont relativement semblables aux distances réelles dans la protéine (donc dans ce cas, la distorsion est minime entre la carte de recombinaison et la carte physique).

Dans la comparaison entre gènes et protéines, nous en sommes réduits à considérer la séquence d'ADN entre les deux points qui codent pour les extrémités de la protéine. Cependant, un gène n'est pas directement traduit en protéine, mais est exprimé via la production d'un **ARN messager**, un acide nucléique intermédiaire

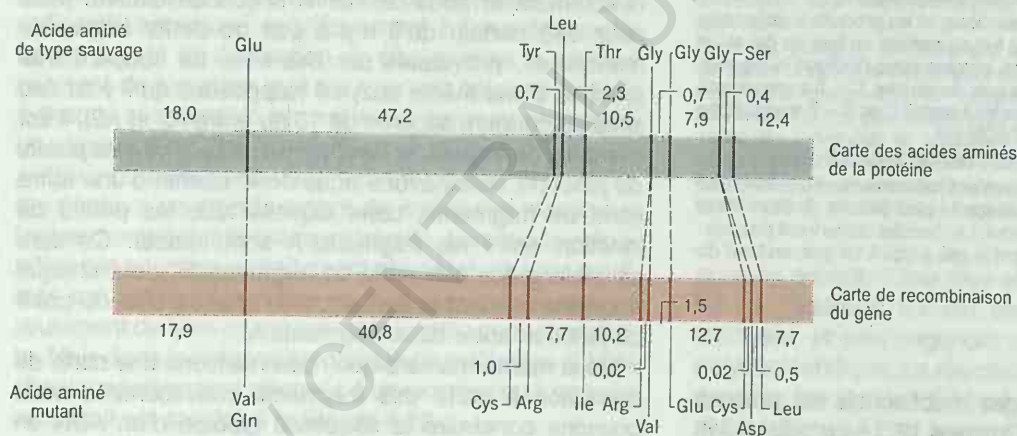


Figure 4.14

La carte de recombinaison du gène de la tryptophane synthétase correspond à la séquence des acides aminés de la protéine.

Les traits noirs indiquent les sites de mutation identifiés soit par des substitutions d'acides aminés dans la séquence protéique, soit par la localisation de mutations sur le gène. La distance entre les barres indique leur pourcentage de séparation sur la carte (donné plus précisément par les chiffres). La carte de recombinaison augmente les distances entre certaines mutations, mais à part cela, correspond bien à la structure physique du gène et de la protéine.

Notons que dans deux cas, des mutations peuvent être séparées sur la carte génétique, mais concernent le même acide aminé sur la carte supérieure (les lignes de connexion convergent). Dans un cas, Gly est changé en Arg ou Val, dans l'autre, Gly est changé en Cys ou Asp. Cela nous indique formellement que l'unité de recombinaison génétique (qui est en réalité d'une paire de bases) est plus petite que l'unité qui code pour un acide aminé (en réalité de trois paires de bases).

en fait utilisé pour synthétiser une protéine (comme nous le verrons en détail dans la Partie 2). L'ARN messager est synthétisé par le même procédé d'appariement des bases complémentaires, utilisé pour répliquer l'ADN, avec la différence majeure qu'il ne correspond qu'à un seul brin de la double hélice d'ADN. Ainsi la séquence d'un ARN messager est complémentaire de la séquence d'un brin d'ADN, et identique (mis à part le remplacement de T par U) à l'autre brin d'ADN.

Un ARN messager comprend une séquence de nucléotides correspondant à la séquence des acides aminés de la protéine. Cette partie de l'acide nucléique est appelée **partie codante**. Mais l'ARN messager peut contenir des séquences supplémentaires à chaque extrémité ; ces séquences ne représentent pas directement la protéine. Cependant, on a l'habitude de considérer que le « gène » contient la séquence entière représentée dans l'ARN messager. Parfois on trouve des mutations qui bloquent la fonction du gène dans des régions supplémentaires non codantes, cela confirme que ces régions font légitimement partie de l'unité génétique.

La **Figure 4.15** illustre la situation où un gène serait composé d'une séquence continue d'ADN, nécessaire à la production d'une protéine spécifique et qui compren-

draît la séquence codant pour la protéine, mais aussi des séquences non codantes de part et d'autre.

LES GÈNES DES EUCARYOTES PEUVENT ÊTRE MORCELÉS

La vision simple du gène a été remise en cause en 1977 par la découverte des **gènes morcelés**. La première preuve de l'existence de ces morcellements vint d'une comparaison entre la structure de l'ADN et de l'ARN messager correspondant. L'ARN messager comporte toujours une séquence nucléotidique qui correspond exactement, selon les règles du code génétique, à la protéine produite. Cependant, *le gène peut comprendre des séquences supplémentaires se situant à l'intérieur de la région codante et interrompant la séquence représentative de la protéine*.

L'incohérence entre la structure de l'ADN et de l'ARN messager est tout à fait commune chez les eucaryotes, et a été trouvée chez des archéobactéries (représentant une autre branche évolutive de la bactérie), mais on ne l'a jamais trouvée chez les eubactéries (dont dérivent les procaryotes les plus courants).

Les séquences d'ADN comprenant un gène morcelé sont divisées en deux catégories décrites à la **Figure**

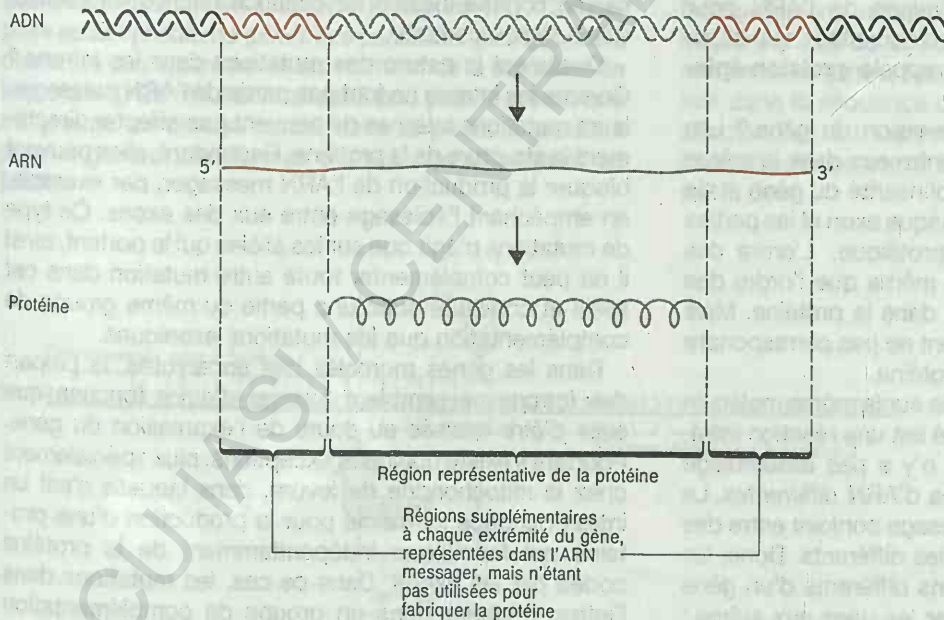


Figure 4.15
Le gène peut être plus long que la séquence codant pour la protéine.

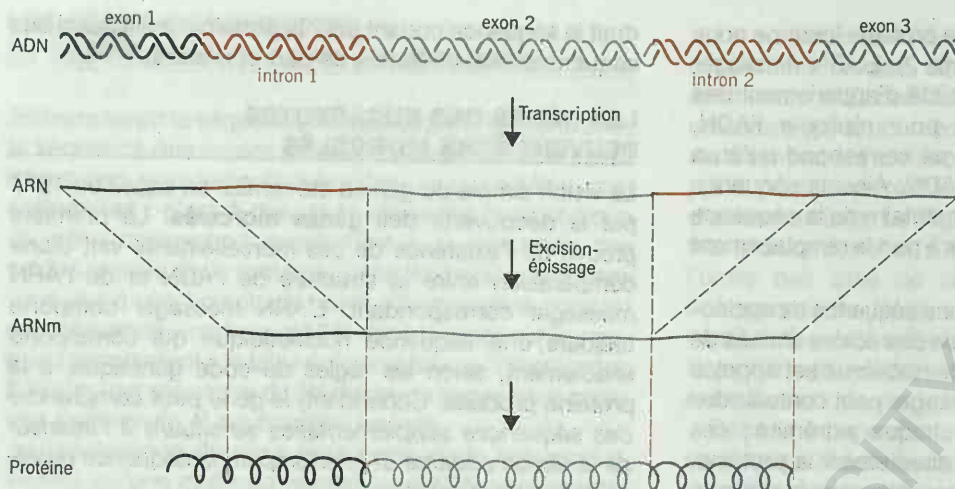


Figure 4.16

Des gènes morcelés sont exprimés via un ARN précurseur, duquel les introns sont excisés et les exons épissés pour donner l'ARN messager.

re 4.16. Les **exons** sont les régions représentées dans l'ARN messager. Les **introns** en sont absents. Le processus d'expression du gène requiert donc une nouvelle étape, absente chez les eubactéries. L'ADN donne naissance à une copie d'ARN qui représente exactement la séquence du génome. Mais cet ARN n'est qu'un précurseur ; il ne peut être utilisé pour produire la protéine. Les introns doivent d'abord être éliminés de l'ARN pour obtenir un ARN messager qui ne soit constitué que d'une série d'exons. Ce processus est appelé **excision-épissage de l'ARN**.

En quoi cela modifie-t-il notre vision du gène ? Les exons sont toujours assemblés entre eux dans le même ordre que dans l'ADN. Ainsi la colinéarité du gène et de la protéine est maintenue entre chaque exon et les parties correspondantes de la chaîne protéique. L'ordre des mutations dans le gène reste le même que l'ordre des substitutions des acides aminés dans la protéine. Mais les *distances* dans le gène peuvent ne pas correspondre du tout aux distances dans la protéine.

Tous les exons sont représentés sur la même molécule d'ARN, et leur épissage simultané est une réaction *intramoléculaire* (voir Chapitre 23). Il n'y a pas assemblage d'exons portés par des molécules d'ARN *différentes*. Le mécanisme exclut donc tout épissage conjoint entre des séquences représentant des allèles différents. Donc, les mutations situées dans des exons différents d'un gène ne peuvent pas se compléter les unes aux autres ; elles continuent donc à être définies comme des membres du même groupe de complémentation.

Si les mutations affectant la séquence d'une protéine d'eucaryote supérieur pouvaient être localisées avec une précision suffisante, on devrait les trouver disposées en groupes. Chaque groupe devrait correspondre à un exon et être séparé du groupe suivant, sur la carte génétique, par une distance correspondant à la longueur de l'intron. Ainsi ne peut-on pas prendre les fréquences de recombinaison comme indices de distance relative à l'intérieur d'une région codante.

Quelle est la nature des mutations dans les introns ? Comme les introns ne font pas partie de l'ARN messager, leurs mutations internes ne peuvent pas affecter directement la structure de la protéine. Cependant, elles peuvent bloquer la production de l'ARN messager, par exemple, en empêchant l'épissage entre eux des exons. Ce type de mutations n'agit que sur les allèles qui le portent, ainsi il ne peut compléter toute autre mutation dans cet allèle et constitue donc une partie du même groupe de complémentation que les mutations exoniques.

Dans les gènes morcelés des eucaryotes, la plupart des introns ne semblent exercer d'autre fonction que celle d'être excisés au cours de l'expression du gène. Pourtant il existe quelques exceptions, plus spécialement chez la mitochondrie de levure, dans laquelle c'est un intron qui code lui-même pour la production d'une protéine qui fonctionne indépendamment de la protéine codée par les exons. Dans ce cas, les mutations dans l'intron tombent dans un groupe de complémentation différent de celui représenté par les mutations dans les exons. Ainsi, un groupe de complémentation peut se

situer entre des ensembles de mutations qui eux-mêmes font partie d'un autre groupe.

Les gènes eucaryotes ne sont pas nécessairement morcelés. Certains correspondent directement au produit de la protéine de la même manière que les gènes eubactériens. Nous ne connaissons pas encore les proportions relatives de gènes morcelés et non morcelés chez les eucaryotes, bien qu'il semble que les premiers soient majoritaires.

CERTAINES SÉQUENCES D'ADN CODENT POUR PLUS D'UNE PROTÉINE

La plupart des gènes sont constitués d'une séquence d'ADN qui n'existe que dans le but unique de coder pour une protéine (bien que le gène puisse contenir des régions non codantes à chaque extrémité, et des introns à l'intérieur des régions codantes). Cependant, dans certains cas, une séquence d'ADN peut coder pour plus d'une protéine.

On peut trouver des gènes *chevauchants* dans la situation relativement simple où un gène fait partie d'un autre gène. La première moitié d'un gène (ou la seconde) peut être utilisée indépendamment pour déterminer une protéine qui représente la première (ou la seconde) moitié d'une protéine elle-même déterminée par le gène entier. Cette inter-relation est illustrée à la **Figure 4.17**. Cela ne pose aucun problème *génétique* particulier (bien que cela demande des ajustements au moment de la synthèse de la protéine). Ce résultat final est à peu près le même que s'il y avait un clivage partiel dans la protéine produite pour engendrer des formes d'une longueur partielle ou totale.

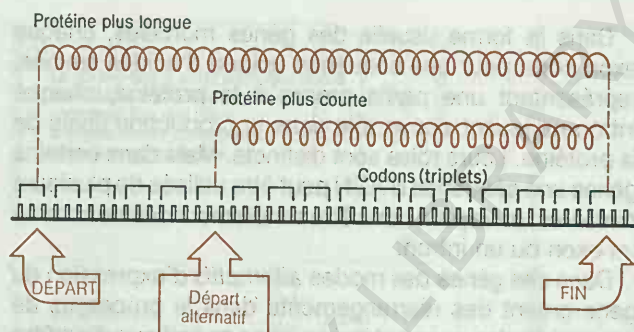


Figure 4.17

Deux protéines peuvent être engendrées par un gène unique dont l'expression commence (ou s'achève) en différents points.

Deux gènes peuvent se chevaucher d'une manière plus subtile quand la même séquence d'ADN est partagée entre deux protéines *non homologues*. Cette situation se produit lorsque la même séquence d'ADN est traduite dans plus d'un cadre de lecture. Dans les gènes cellulaires, une séquence d'ADN est habituellement lue dans un seul des trois cadres potentiels de lecture, mais dans certains gènes viraux et mitochondriaux, deux gènes contigus qui sont lus selon des cadres de lecture différents peuvent se chevaucher. Cette situation est illustrée par la **Figure 4.18**. La distance de recouvrement est habituellement relativement courte, de sorte que la majeure partie de la séquence représentant la protéine ne conserve qu'une fonction codante unique. Une mutation dans la séquence commune peut, selon son genre, affecter l'un ou l'autre gène ou les deux ; ainsi, elle pourrait soit appartenir à un groupe de complémentation, soit à deux.

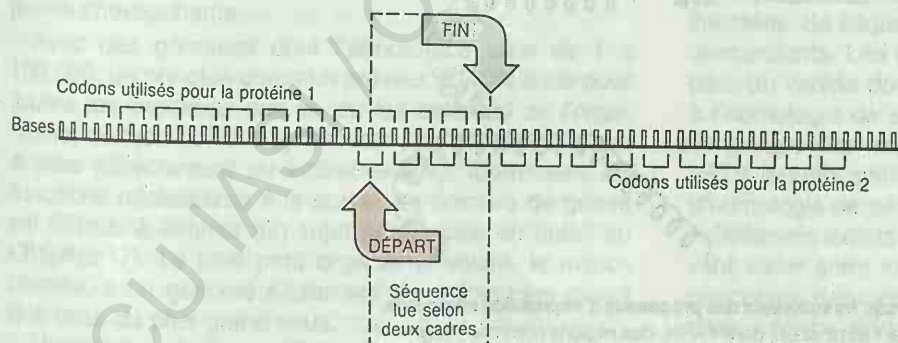


Figure 4.18

Deux gènes peuvent partager la même séquence si l'ADN est lu selon des cadres différents.

Dans la forme usuelle des gènes morcelés, chaque exon code pour une séquence unique d'acides aminés, représentant une partie propre à la protéine, chaque intron ne jouant aucun rôle dans la production finale de la protéine. Leurs rôles sont distincts. Mais dans certains gènes une séquence d'ADN peut être utilisée de plusieurs manières, et ne peut donc être simplement définie comme un exon ou un intron.

Dans ces gènes des modes *alternatifs* d'expression du gène créent des réarrangements dans le processus de connexion des exons. Ainsi un exon spécifique peut être connecté avec n'importe quel exon alternatif pour former un ARN messager. Les formes alternatives produisent des protéines dans lesquelles une partie est commune mais pas l'autre. La **Figure 4.19** illustre l'exemple de certaines régions qui peuvent se comporter comme des exons quand elles sont exprimées d'une certaine manière, et comme des introns quand elles sont exprimées d'une autre manière.

Parfois les deux processus sont simultanés, une certaine quantité d'ARN étant épissée selon l'un ou l'autre mode, d'autres fois les processus sont alternatifs, c'est-à-dire exprimés dans des conditions différentes, l'un dans un certain type cellulaire et l'autre dans un autre type.

Les propriétés de complémentation de telles mutations sont en principe complexes, et il n'est donc pas forcément possible de rattacher toutes les mutations à un groupe indépendant de complémentation ou à un autre.

L'INFORMATION GÉNÉTIQUE PEUT ÊTRE FOURNIE PAR L'ADN OU PAR L'ARN

Les acides nucléiques obéissent aux critères requis pour un matériel génétique grâce à la combinaison toute particulière de stabilité et de flexibilité. La réplication selon le mécanisme d'appariement des bases complémentaires permet à l'information génétique d'être parfaitement transmise. Pourtant, lorsque surviennent des mutations occasionnelles, des modifications transmissibles se produisent, servant de substrat à l'évolution. Le code génétique nous fournit l'expression de l'information génétique sous la forme de protéines.

Ces mécanismes sont efficaces pour l'information génétique cellulaire des procaryotes aussi bien que des eucaryotes ainsi que pour l'information véhiculée par les virus. Les génomes de tous les organismes vivants sont composés d'ADN en duplex. Les génomes des virus peuvent être constitués d'ADN ou d'ARN et il y existe

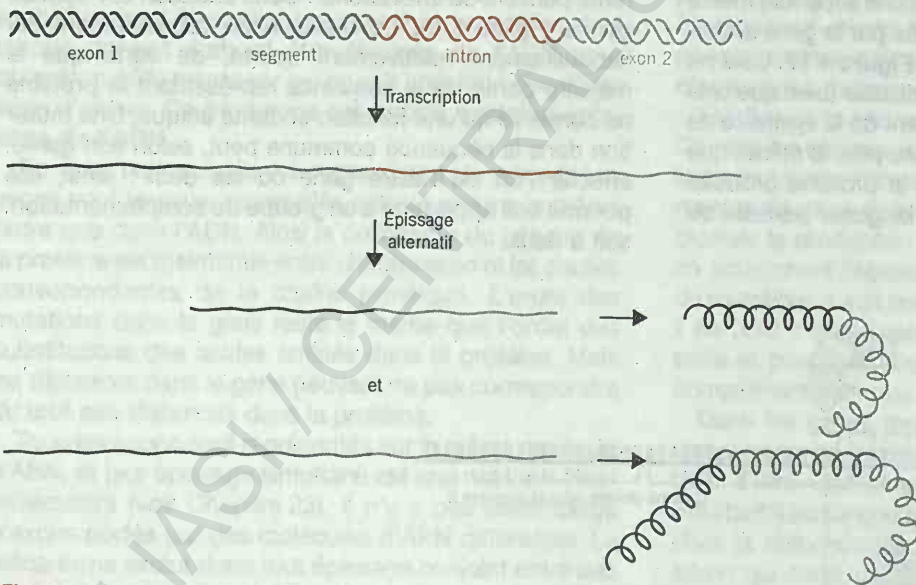


Figure 4.19

Les gènes peuvent être difficiles à délimiter lorsqu'ils subissent des processus d'expression alternatifs.

Dans un processus, l'*exon 1* se lie à l'*exon 2* par l'élimination, dans l'ARN, des régions nommées *segment* et *intron*. Dans l'autre processus, la région de l'*exon 1*-*segment* se lie directement à l'*exon 2*, ne faisant disparaître que l'*intron*. Ainsi, dans le premier schéma, la région *segment* est un intron, alors que dans l'autre, c'est un exon. Les processus produisent deux protéines dont les extrémités sont semblables, mais dont l'une a une séquence supplémentaire en son milieu. Ainsi une région d'ADN peut coder pour plus d'une seule protéine.

des exemples de chaque type qui sont soit double-brin (ds = double strand), soit simple-brin (ss = single strand). Des détails du mécanisme utilisé pour répliquer l'acide nucléique peuvent varier selon les systèmes viraux, mais le principe de réplication par synthèse de brins complémentaires demeure le même, tel que le montre la **Figure 4.20**.

Certains génomes de virus double-brin sont constitués par de l'ADN et d'autres par de l'ARN. Les deux peuvent être répliqués par le mécanisme de réplication semi-conservative, en utilisant les brins séparés comme matrice pour la synthèse des partenaires.

Les virus à génomes simple-brin utilisent ce simple-brin comme matrice pour synthétiser un brin complémentaire ; et celui-ci est utilisé, à son tour, pour synthétiser son complément, qui est, bien sûr, identique au brin originel du départ. La réplication peut impliquer la formation d'intermédiaires double-brins stables, ou peut n'utiliser l'acide nucléique double-brins que comme une étape transitoire.

Les principes sont donc les mêmes pour perpétuer l'information génétique, que ce soit pour les énormes génomes des plantes ou des amphibiens, pour les minuscules génomes du mycoplasme, ou l'information génétique encore plus petite des virus à ADN ou à ARN. Le **Tableau 4.2** résume quelques exemples qui illustrent l'étendue des variétés et des tailles des génomes.

Les génomes à ADN de toutes les cellules vivantes sont constitués de très longues molécules ; tous, sauf la bactérie et le mycoplasme, possèdent plus d'une telle molécule. Certains des plus grands virus (comme l'influenza par exemple) ont des génomes segmentés constitués de plus d'une molécule d'acide nucléique. D'autres génomes viraux sont constitués d'une seule molécule d'acide nucléique. Pour les virus les plus petits (SV40, ϕ X, MS2), les différentes fonctions peuvent impliquer des gènes chevauchants.

Avec des génomes dont l'abondance varie de 1 à 100 000, un principe commun prévaut. *L'ADN code pour toutes les protéines que la, ou les cellule(s) de l'organisme, doit (ou doivent) synthétiser ; les protéines, quant à elles (directement ou indirectement), fournissent les fonctions nécessaires à la survie.* Le nombre de gènes est difficile à estimer (un sujet étudié plus en détail au Chapitre 17). Le plus petit organisme vivant, le mycoplasme, a un génome seulement deux fois plus grand que celui du plus grand virus.

Un principe similaire définit la fonction de l'information génétique des virus qu'ils soient à ADN ou à ARN. *L'acide nucléique code pour la ou les protéine(s) nécessaire(s) à l'empaquetage (in vivo) du génome et également pour*

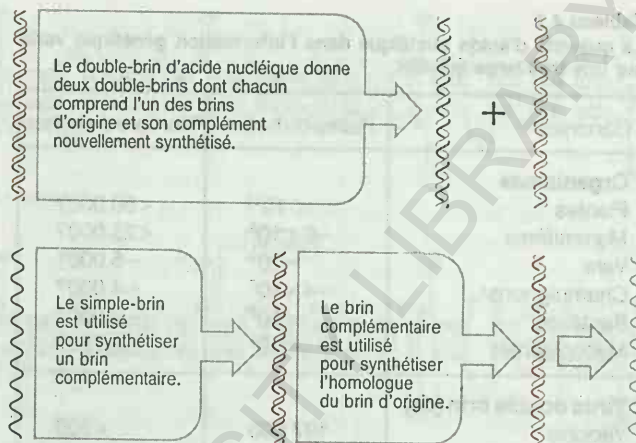


Figure 4.20

Les acides nucléiques simple ou double-brin se répliquent tous deux par synthèse des brins complémentaires selon les règles d'appariement des bases.

toutes les fonctions qui, outre celles de la cellule hôte, sont nécessaires à la reproduction du virus pendant son cycle infectieux. (Le plus petit virus, le virus satellite de la nécrose du tabac — STNV* —, ne peut se répliquer seul, mais a besoin de la présence simultanée d'un virus « aide » — helper — : le virus de la nécrose du tabac, TNV* qui est lui-même un virus normalement infectieux.)

Les **viroïdes** sont une nouvelle espèce d'agents **pathogènes subviraux**, dont les principes de survie peuvent être différents. Les viroïdes sont des agents infectieux qui causent des maladies des plantes supérieures. Ce sont de très petites molécules circulaires d'ARN. Contrairement aux virus, dont l'agent infectieux est un *virion* contenant un génome encapsulé dans une enveloppe protéique, *l'ARN viroïde est par lui-même l'agent infectieux*.

Un ARN viroïde est une espèce moléculaire unique qui se réplique de manière autonome dans les cellules infectées. Sa séquence se perpétue fidèlement chez ses descendants. Les viroïdes rentrent dans plusieurs groupes. Un viroïde donné se classe dans un groupe, grâce à l'homologie de sa séquence avec d'autres membres du groupe. Par exemple, quatre viroïdes apparentés au PSTV (potato spindle tuber viroid) présentent 70-83 % d'homologie de séquence avec lui.

Différents isolats d'une souche viroïde spécifique peuvent varier entre eux, et la modification peut affecter le phénotype des cellules infectées. Par exemple, les souches « mild » (*modéré*) et « severe » (*grave*) du PSTV diffèrent par trois substitutions de nucléotides.

Ces propriétés pourraient aussi bien caractériser des

* STNV : satellite tobacco necrosis virus.

TNV : tobacco necrosis virus.

Tableau 4.2

La quantité d'acide nucléique dans l'information génétique varie sur une très large échelle.

Génomes	Paires de bases	Nombre de gènes
Organismes		
Plantes	$< 10^{11}$	$< 50\ 000?$
Mammifères	$\sim 3 \times 10^8$	$< 25\ 000?$
Vers	$\sim 10^8$	$\sim 5\ 000?$
Champignons	$\sim 4 \times 10^7$	$\sim 4\ 000?$
Bactéries	$< 10^7$	$\sim 2\ 000?$
Mycoplasmes	$< 10^6$	$\sim 750?$
Virus double brin (ds)		
Vaccine	187 000	< 300
Papovavirus (SV40)	5 226	~ 6
Bactériophage T4	165 000	~ 200
Virus simple brin (ss)		
Parvovirus	5 000	5
Bactériophage ϕ X174	5 387	11
Virus à ARN double brin (ds)		
Reovirus	23 000	22
Virus à ARN simple brin		
Coronavirus	20 000	7
Influenza	13 500	12
TMV	6 400	4
Bactériophage MS2	3 569	4
STNV	1 300	1
Viroïdes		
PSTV à ARN	359	—
Scrapie		
Prion de la scrapie	—	—

virus conventionnels. Comme les virus, les viroïdes possèdent des acides nucléiques dont les séquences peuvent se répliquer de façon conservative, sont sujettes à mutations et déterminent le phénotype. Là encore, ils remplissent les critères de l'information génétique. Pourtant les viroïdes diffèrent des virus par leur structure et leur fonction.

Quant à sa structure, l'agent viroïde n'est inséré dans aucun environnement protecteur. Il est constitué uniquement d'ARN, apparié sur une grande longueur mais imparfaitement, et formant une baguette linéaire caractéristique comme le montre l'exemple de la **Figure 4.21**.

Le viroïde à ARN ne semble pas transcrit en protéine,

et de ce fait il ne peut pas lui-même coder pour les fonctions nécessaires à sa survie. Cette situation pose deux questions. Comment le viroïde à ARN peut-il se répliquer ? Et comment affecte-t-il le phénotype de la cellule infectée de la plante ?

La réplication doit être réalisée par les enzymes de la cellule hôte, détournées de leur fonction normale. Le fait que la séquence viroïde soit héréditaire indique que ce viroïde à ARN en constitue la matrice.

Les viroïdes sont probablement pathogènes car ils interfèrent avec les processus cellulaires normaux. Il se pourrait qu'ils le fassent relativement au hasard, par exemple, en bloquant une enzyme essentielle à leur propre réplication, ou en interférant avec la production des ARN cellulaires nécessaires. Alternativement, ils pourraient représenter des molécules régulatrices anormales, ayant des effets particuliers sur l'expression des gènes individuels.

La **scrapie** est un agent encore plus rare, il cause une maladie dégénérative neurologique chez le mouton et la chèvre. Cette maladie peut être rapprochée des maladies humaines du kourou et du syndrome de Creutzfeld-Jacob qui affecte les fonctions du cerveau.

L'agent infectieux de la scrapie ne semble contenir aucun acide nucléique. Des tentatives répétées pour identifier ne serait-ce qu'un petit ADN ou ARN ont échoué. Le seul composant ayant été jusqu'à présent identifié, est une protéine extrêmement résistante aux attaques par les protéases.

Cet agent exceptionnel est parfois décrit comme un **prion**. Son composant majeur est une glycoprotéine hydrophobe de 28 000 daltons, la PrP, qui peut s'agréger en structures en bâtonnets semblables à celles que l'on trouve dans le cerveau de personnes atteintes par la maladie d'Alzheimer, et d'autres syndromes neurologiques.

La PrP est codée par un gène cellulaire (qui est conservé chez les mammifères) et qui est exprimé dans le cerveau normal. Le produit cellulaire est différent du prion : il est totalement détruit par les protéases, ce qui contraste avec la résistance de la PrP. Nous supposons qu'une modification lui conférant une résistance aux protéases est impliquée dans la formation de la PrP.

Si la PrP est effectivement l'agent infectieux de la scrapie, il doit, d'une certaine manière, modifier la synthèse de son homologue cellulaire normal afin de le rendre infectieux alors qu'il ne l'était pas. Un tel arrangement satisferait aux restrictions formelles apportées au dogme central, mais cette idée serait en désaccord avec l'esprit du paradigme.

Une autre hypothèse serait que les préparations de

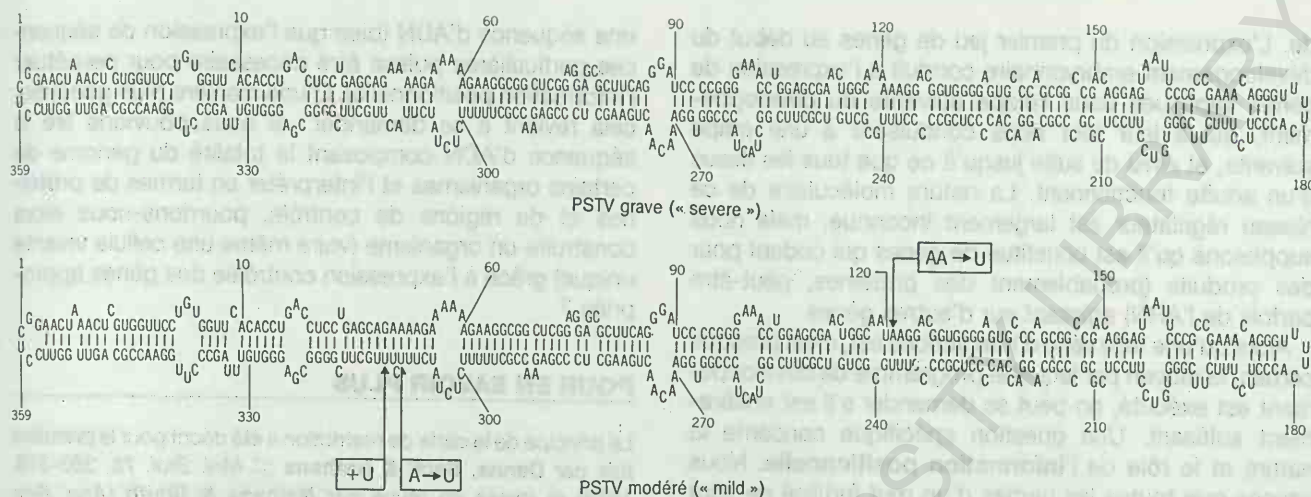


Figure 4.21

L'ARN PSTV est une molécule circulaire qui forme une longue structure double-brin, interrompue par de nombreuses boucles intérieures. Les formes « severe » et « mild » diffèrent en trois emplacements.

prions contiendraient des quantités extrêmement faibles d'un acide nucléique relativement petit et extrêmement infectieux. Pour être indétectable dans des conditions expérimentales, l'acide nucléique (virus ou viroïde) devrait être plus infectieux que les agents déjà connus.

LA PORTÉE DU PARADIGME

Les catégories exceptionnelles d'agents, que sont les viroïdes et la scrapie, démontrent que l'information héréditaire peut être transmise par des formes autres que celles d'un gène codant pour une protéine. Pourtant, dans chaque cas, nous avons été réduits à analyser le devenir de l'agent selon son interaction ou son interférence avec des gènes cellulaires.

Le concept du gène lui-même a néanmoins récemment évolué. La question de ce que recouvre une désignation s'applique particulièrement au gène. Il est clair que nous ne pouvons plus dire qu'un gène est une séquence d'ADN qui code uniquement et de façon continue pour une protéine particulière. Dans les cas où une séquence d'ADN est responsable de la production d'une protéine donnée, l'usage actuel est de considérer que la séquence entière d'ADN du premier point représenté dans l'ARN messager au dernier point correspondant à son extrémité, délimite le gène, exons, introns compris.

Lorsque les séquences représentant les protéines se chevauchent, ou ont des formes d'expression alternati-

ves, on peut inverser la définition habituelle du gène. Au lieu de dire « un gène-un polypeptide », on peut définir la relation : « un polypeptide-un gène ». Ainsi on peut considérer la séquence réellement responsable de la production du polypeptide (introns comme exons) comme le gène, tout en reconnaissant que dans la perspective d'une autre protéine, une partie de cette même séquence pourrait également appartenir à son gène. Ceci nous permet des descriptions telles que celles des gènes « chevauchants » ou « alternatifs ».

Nous pouvons donc maintenant voir combien nous avons progressé depuis l'hypothèse originelle de Beadle et Tatum. A cette époque l'idée directrice était la nature du gène. Une fois découvert que les gènes représentaient des protéines, le paradigme a été établi sous la forme du concept : chaque unité génétique fonctionne par la synthèse d'une protéine particulière.

Cette vision demeure le paradigme central de la biologie moléculaire : une séquence d'ADN fonctionne soit en codant directement pour une protéine donnée, soit en étant nécessaire à l'utilisation d'un segment adjacent codant réellement pour la protéine. Par delà l'explication des relations de base entre les gènes et les protéines, jusqu'où ce paradigme nous entraîne-t-il ?

Le développement des organismes multicellulaires repose sur l'utilisation de différents gènes pour engendrer les différents phénotypes de la cellule dans chaque tissu. L'expression des gènes est déterminée par un réseau régulateur qui prend probablement la forme d'une casca-

de. L'expression du premier jeu de gènes au début du développement embryonnaire conduit à l'expression de gènes impliqués dans l'étape suivante du développement, qui à leur tour nous conduisent à une étape suivante, et ainsi de suite jusqu'à ce que tous les tissus d'un adulte fonctionnent. La nature moléculaire de ce réseau régulateur est largement inconnue, mais nous supposons qu'il est constitué de gènes qui codent pour des produits (probablement des protéines, peut-être parfois de l'ARN) agissant sur d'autres gènes.

Alors qu'une telle série d'interactions est, c'est presque certain, le moyen par lequel le programme de développement est exécuté, on peut se demander s'il est entièrement suffisant. Une question spécifique concerne la nature et le rôle de l'**information positionnelle**. Nous savons que toutes les parties d'un œuf fertilisé ne sont pas équivalentes ; une des caractéristiques responsables du développement de différentes parties du tissu à partir de régions différentes de l'œuf est la position de l'information (probablement des macromolécules spécifiques) à l'intérieur de la cellule.

Nous ne savons pas comment ces régions spécifiques se forment. Mais nous pouvons spéculer que l'existence de l'information positionnelle dans l'œuf même conduit à l'expression différentielle de gènes dans les cellules formées ultérieurement dans ces régions, ce qui conduit au développement de l'organisme adulte, qui conduit au développement d'un œuf possédant l'information positionnelle appropriée...

Cette possibilité nous conduit à nous demander si une partie de l'information nécessaire au développement de l'organisme pourrait n'être pas directement attribuable à

une séquence d'ADN (bien que l'expression de séquences particulières puisse être nécessaire pour perpétuer l'information positionnelle). D'une manière plus générale, cela revient à se demander : si nous pouvions lire la séquence d'ADN composant la totalité du génome de certains organismes et l'interpréter en termes de protéines et de régions de contrôle, pourrions-nous alors construire un organisme (voire même une cellule vivante unique) grâce à l'expression contrôlée des gènes appropriés ?

POUR EN SAVOIR PLUS

Le principe de la carte de restriction a été décrit pour la première fois par **Danna, Sack & Nathans** (*J. Mol. Biol.* **78**, 363-376, 1973) et passé en revue par **Nathans & Smith** (*Ann. Rev. Biochem.* **46**, 273-293, 1975). La construction des cartes de liaison d'après des polymorphismes de restriction a été explorée par **White et al.** (*Nature*, **313**, 101-105, 1985). Les méthodes de séquençage de l'ADN ont été passées en revue par **Wu** (*Ann. Rev. Biochem.* **47**, 607-734, 1978). Les deux méthodes les plus populaires sont celles de **Maxam & Gilbert** (*Proc. Nat. Acad. Sci. USA* **74**, 560-564, 1977) et de **Sanger, Nicklen & Coulson** (*Proc. Nat. Acad. Sci. USA*, **74**, 5463-5467, 1977). Les formes que les gènes peuvent prendre et les implications génétiques de leur organisation sont discutées dans **Lewin's, Gene Expression, 2, Eukaryotic Chromosomes** (Wiley, New-York, 1980). Les viroïdes ont été analysés par **Diener** (*Adv. Virus Res.* **28**, 241-283, 1983 ; *Proc. Nat. Acad. Sci. USA*, **83**, 58-62, 1986). Les prions de la scrapie ont été identifiés à la protéine PrP par **Mc Kinley, Bolton & Prusiner** (*Cell* **35**, 57-62, 1983), et leurs gènes identifiés par **Oesch et al.** (*Cell* **40**, 735-746, 1985).

PARTIE 2

TRANSFORMER LES GÈNES EN PROTÉINES

Le concept de base établit que « l'information », une fois introduite dans la protéine, ne peut plus en ressortir. Le transfert de l'information d'un acide nucléique à un autre acide nucléique, ou d'un acide nucléique à une protéine peut être possible, mais le transfert d'une protéine à une protéine, ou d'une protéine à un acide nucléique est impossible. Ici, l'information signifie la détermination précise d'une séquence, qu'elle soit constituée de bases dans l'acide nucléique ou d'acides aminés dans la protéine.

Francis Crick, 1958

the first of these is the fact that the system of the
 the second is the fact that the system of the
 the third is the fact that the system of the
 the fourth is the fact that the system of the
 the fifth is the fact that the system of the

the sixth is the fact that the system of the
 the seventh is the fact that the system of the
 the eighth is the fact that the system of the
 the ninth is the fact that the system of the
 the tenth is the fact that the system of the

the eleventh is the fact that the system of the
 the twelfth is the fact that the system of the
 the thirteenth is the fact that the system of the
 the fourteenth is the fact that the system of the
 the fifteenth is the fact that the system of the

the sixteenth is the fact that the system of the
 the seventeenth is the fact that the system of the
 the eighteenth is the fact that the system of the
 the nineteenth is the fact that the system of the
 the twentieth is the fact that the system of the

the twenty-first is the fact that the system of the
 the twenty-second is the fact that the system of the
 the twenty-third is the fact that the system of the
 the twenty-fourth is the fact that the system of the
 the twenty-fifth is the fact that the system of the

the twenty-sixth is the fact that the system of the
 the twenty-seventh is the fact that the system of the
 the twenty-eighth is the fact that the system of the
 the twenty-ninth is the fact that the system of the
 the thirtieth is the fact that the system of the

the thirty-first is the fact that the system of the
 the thirty-second is the fact that the system of the
 the thirty-third is the fact that the system of the
 the thirty-fourth is the fact that the system of the
 the thirty-fifth is the fact that the system of the

the thirty-sixth is the fact that the system of the
 the thirty-seventh is the fact that the system of the
 the thirty-eighth is the fact that the system of the
 the thirty-ninth is the fact that the system of the
 the fortieth is the fact that the system of the

BCU IASI / CENTRAL UNIVERSITY LIBRARY

CHAPITRE 5

LA CHAÎNE DE MONTAGE DE LA SYNTHÈSE PROTÉIQUE

Le **concept central** définit le paradigme de la biologie moléculaire : les gènes sont des unités qui se perpétuent et qui fonctionnent à travers leur expression sous forme de protéines.

L'information génétique est portée par la séquence de l'ADN. Elle est conservée par la **réplication** quand deux copies identiques proviennent de la duplication de l'acide nucléique. L'information est exprimée grâce à un mécanisme en deux étapes. La **transcription** engendre un ARN simple brin dont la séquence est identique à l'un des deux brins de l'ADN bicaténaire. Puis, la **traduction** convertit la séquence nucléotidique de l'ARN en une séquence d'acides aminés constituant une protéine.

La découverte du code génétique a montré que l'information génétique est conservée sous forme de triplets nucléotidiques, sans avoir révélé la spécificité de chaque codon vis-à-vis de son acide aminé correspondant.

Le concept de la nécessité absolue d'un code évoluait de pair avec l'idée que la traduction impliquait une **matrice**. Puisque dans une cellule eucaryote le matériel génétique, nucléaire, est physiquement séparé du site de synthèse cytoplasmique, il était évident que l'ADN en lui-même ne pouvait pas servir de matrice.

Celle-ci est élaborée par transcription, sous la forme d'un **ARN messager (ARNm en abrégé)** qui représente

un brin d'ADN. Le brin d'ADN qui porte la *même* séquence que l'ARNm (exception faite de T remplacé par U) est appelé **brin codant**. L'autre brin d'ADN qui dirige la synthèse d'ARNm via l'appariement de bases complémentaires est le **brin anticodant**. Puisque le code génétique est réellement *lu* sur l'ARNm, il est habituellement représenté par les quatre bases de l'ARN : U, C, A, G.

Décrire l'ARN *messenger* comme matrice reflète bien sa capacité (chez les eucaryotes) à passer du noyau où il est synthétisé, au cytoplasme où il est opérationnel. La traduction de l'ARNm en protéine s'accomplit par la lecture du code génétique : chaque triplet de nucléotides est converti en un acide aminé. Donc, la traduction constitue l'étape au cours de laquelle la séquence nucléotidique est décodée en acides aminés.

La **Figure 5.1** illustre les rôles de la réplication, de la transcription et de la traduction dans l'optique du concept central.

Le *matériel génétique se perpétue sous forme d'ADN ou d'ARN* (voir Figure 4.20). Les cellules n'utilisent que l'ADN. Certains virus utilisent l'ARN dont la réplication est possible dans la cellule infectée.

L'*expression de l'information génétique cellulaire est normalement irréversible*. La transcription de l'ADN produit des molécules d'ARN dont l'*unique fonction* sera de

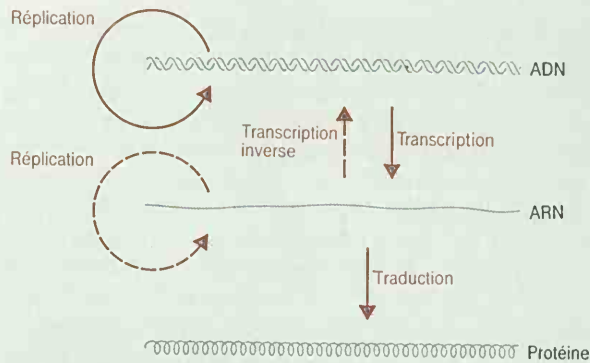


Figure 5.1

Le concept de base établit que l'information contenue dans l'acide nucléique peut être perpétuée ou transférée mais le transfert de l'information dans la protéine est irréversible.

Les mécanismes cellulaires sont indiqués en lignes continues, tandis que ceux utilisés uniquement dans l'infection virale le sont en pointillés.

produire ultérieurement une séquence protéique ; en règle générale, elles ne seront pas réutilisées comme support d'information génétique.

La restriction à une voie unidirectionnelle allant de l'ADN à l'ARN n'est pas une règle absolue. Elle est transgressée par les **rétrovirus**, dont le génome est constitué par un ARN monocaténaire. Au cours du cycle d'infection, l'ARN est converti en ADN simple brin par **transcription inverse** puis en ADN bicaténaire. Ce dernier peut alors s'intégrer au génome de la cellule et se répliquer comme les autres gènes. Ainsi la *transcription inverse permet le recouvrement d'une séquence d'ARN et son utilisation en tant qu'information génétique*.

L'existence de la réplication et de la transcription inverse de l'ARN établit le principe général : *l'information contenue dans la séquence d'un type d'acide nucléique peut être convertie en l'autre type d'acide nucléique*. Mais habituellement, aucun de ces mécanismes n'est utilisé par la cellule elle-même qui emploie uniquement les processus de réplication, transcription et traduction. Cependant, en de rares occasions (sans doute induites par un ARN viral), l'information contenue dans un ARN cellulaire peut être récupérée et insérée dans le génome. Bien que la transcription inverse n'ait aucun effet sur les opérations régulières de la cellule, elle devient cependant un mécanisme potentiellement important quand on considère l'évolution du génome.

L'ARN DE TRANSFERT EST UN ADAPTATEUR

La disparité de structure entre une séquence de trinucléotides et un acide aminé soulève d'emblée le problème de la correspondance entre chaque codon et son acide aminé spécifique. Avant même que la structure exacte du code ait été élucidée, Crick suggérait que la traduction nécessitait l'intermédiaire d'un élément « adaptateur ». Cet adaptateur est l'**ARN de transfert (ARNt en abrégé)**, une petite molécule dont la chaîne polynucléotidique est constituée de 75-85 bases. Il existe au moins un ARNt pour chaque acide aminé.

(Par convention, un ARNt est représenté par l'abréviation en trois lettres de l'acide aminé en indice supérieur ; s'il y a plus d'un ARNt pour le même acide aminé, on utilise pour les distinguer des numéros en indice inférieur. Par exemple ARNt^{Ala} est un ARNt pour l'alanine.)

Un ARNt possède deux propriétés fondamentales. Premièrement, il ne peut reconnaître qu'un seul acide aminé auquel il est *lié de façon covalente*. Deuxièmement, chaque ARNt contient une séquence de trinucléotides, l'**anticodon complémentaire du codon qui représente son acide aminé**. L'anticodon permet à l'ARNt de reconnaître le codon via l'appariement de bases complémentaires.

La structure secondaire de l'ARN de transfert est caractéristique. La séquence nucléotidique de tous les ARNt peut être représentée sous la forme d'une **feuille de trèfle (Figure 5.2)** dans laquelle les branches formées de bases complémentaires appariées se terminent par des **boucles simple-brin**. Les structures branches-boucles forment les **bras** de l'ARNt.

On désigne par **aminoacyl-ARNt**, l'ARNt qui a pris en **charge** l'acide aminé qui correspond à son anticodon. Une liaison ester relie le carboxyle de l'acide aminé à l'hydroxyle du ribose de la dernière base de l'ARNt (cette base est toujours l'adénine).

Le mécanisme de liaison entre l'acide aminé et l'ARNt est catalysé par une enzyme spécifique, l'**aminoacyl-ARNt synthétase**. Il existe (au moins) 20 aminoacyl-ARNt synthétases. Chacune ne reconnaît qu'un seul acide aminé et tous les ARNt sur lesquels il peut légitimement être situé.

Est-ce uniquement la séquence de l'anticodon qui permet à l'aminoacyl-ARNt synthétase de reconnaître le codon approprié ? La **Figure 5.3** montre une expérience qui répond à cette question. Une désulfuration oxydative transforme l'acide aminé du cystéinyl-ARNt en alanine, créant un alanyl-ARNt^{Cys}. Le résidu alanine est incorporé dans la protéine au lieu de la cystéine. Une fois l'ARNt

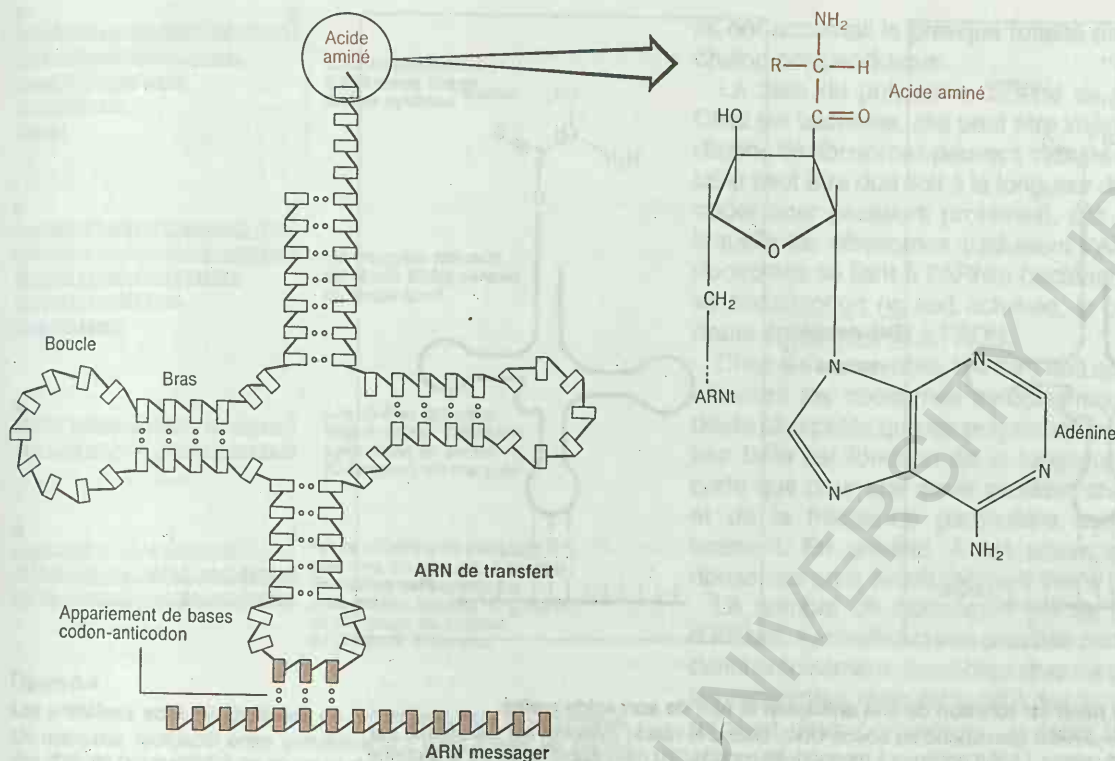


Figure 5.2
L'ARN de transfert a les propriétés doubles d'un adaptateur.

Les ensembles sucre-base sont représentés par des rectangles : les lignes qui les relient sont les liaisons phosphodiester. Les tiges de chaque bras sont maintenues par l'appariement de bases complémentaires (indiquées par de petits cercles) ; les bases à l'extrémité de chaque bras forment une boucle simple-brin. Sur la boucle inférieure, une séquence anticodon est complémentaire du codon pour l'acide aminé que représente l'ARNt.

Sur le bras supérieur, l'acide aminé est lié au nucléotide en 3' de l'ARNt. L'aminocyl-ARNt est constitué d'un acide aminé lié par une liaison ester soit au 2'-OH soit au 3'-OH terminal de l'ARNt qui se termine toujours par la séquence de trois bases -CCA.

chargé, l'acide aminé ne joue plus d'autre rôle dans sa spécificité qui est déterminée exclusivement par l'anticodon.

L'ARN MESSAGER EST TRADUIT PAR LES RIBOSOMES

Si le code génétique se lit comme une suite de triplets adjacents, la protéine qui lui correspond doit être assemblée séquentiellement d'une extrémité à une autre. Dans quelle direction s'effectue ce mécanisme ?

Cette question fut résolue en 1961, quand Dintzis

étudia la synthèse de l'hémoglobine, protéine des globules rouges sanguins. Dans les cellules synthétisant l'hémoglobine, l'un des acides aminés fut remplacé par son équivalent marqué. Puis, les chaînes protéiques achevées furent extraites et la radioactivité de toutes les parties de la chaîne mesurée.

La Figure 5.4 montre qu'initialement la radioactivité s'intègre dans les chaînes polypeptidiques dans la position qui correspond au degré de synthèse que chaque molécule a atteint individuellement. Mais, dans les chaînes achevées, la radioactivité apparaît d'abord dans l'extrémité qui est synthétisée *en dernier*. Après un temps de marquage plus long, la radioactivité est détectée dans

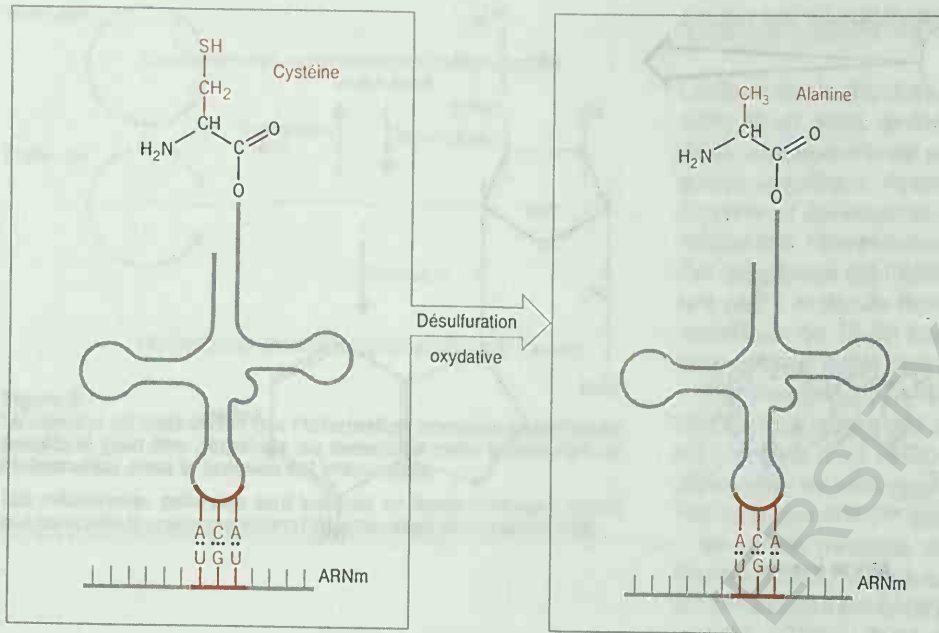


Figure 5.3

La signification d'un ARNt est fonction de son anticodon et non de son acide aminé.

L'anticodon du cystéinyl-ARNt correspond au codon UGU. Quand le résidu cystéinyl est transformé par voie chimique en résidu alanyl, l'ARNt continue à répondre au codon UGU mais apporte alors à la protéine une alanine au lieu d'une cystéine.

toutes les parties de la chaîne synthétisées antérieurement.

Donc, la radioactivité agit en reculant, de la fin vers le début. Cette expérience a montré que *l'assemblage d'une chaîne protéique ne s'effectue que dans une seule direction, du N-terminal vers le C-terminal*.

Les acides aminés sont assemblés en protéines par le **ribosome**, une **particule ribonucléoprotéique** compacte, formée de deux sous-unités. Chaque sous-unité est constituée de plusieurs protéines associées à une longue molécule d'ARN ; ces ARN sont les **ARN ribosomaux** (ARNr en abrégé).

On pensait, quand les ribosomes ont été découverts, qu'ils pourraient réellement *fournir* la matrice de la synthèse protéique. Ceci sous-entendait que différents ribosomes pouvaient contenir différents ARN servant à la synthèse de différentes protéines. C'est en confondant ARNr et ARNm, que cette hypothèse presque correcte manquait de peu son but. Aujourd'hui, nous savons que

tous les ribosomes sont semblables ; simplement ils entreprennent la synthèse de différentes protéines en s'associant aux divers ARNm qui sont, eux, les vraies matrices.

Le ribosome apporte l'environnement qui contrôle la reconnaissance entre un codon de l'ARNm et l'anticodon de l'ARNt. Pour réaliser cette synthèse protéique séquentielle, *le ribosome se déplace le long de l'ARNm*.

Un ribosome se lie à l'extrémité 5' de l'ARNm ; en se déplaçant le long de l'ARN vers l'extrémité 3', il traduit chemin faisant chaque codon en un acide aminé. Au cours de la progression du ribosome, les aminoacyl-ARNt appropriés s'unissent à lui, apportant leurs acides aminés à la chaîne polypeptidique. Le ribosome, continuellement, apparie les deux espèces correspondant aux codons successifs, permettant ainsi la formation d'une liaison peptidique entre deux acides aminés. La chaîne polypeptidique s'allonge d'un acide aminé à chaque étape.

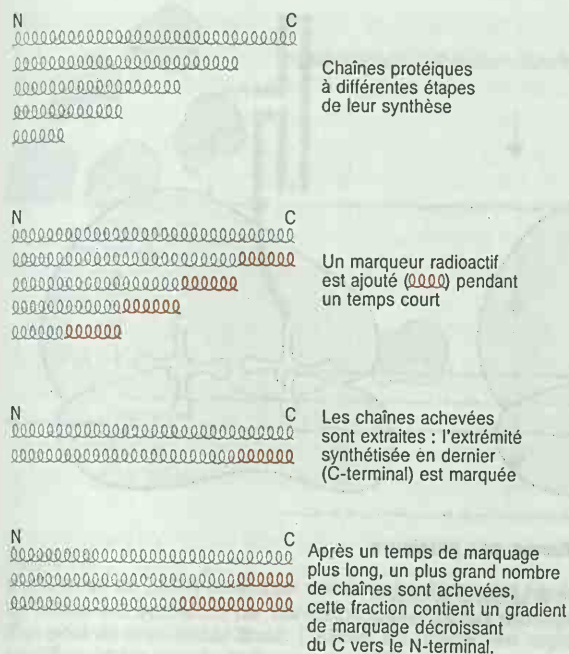


Figure 5.4

Les protéines sont synthétisées du N-terminal vers le C-terminal. Un marqueur radioactif entre simultanément dans toutes les parties des chaînes polypeptidiques en cours de synthèse. Quand les chaînes achevées sont extraites, elles contiennent un marqueur dans la dernière région achevée. Ainsi le gradient de radioactivité du C-terminal vers le N-terminal indique que la synthèse s'effectue dans le sens opposé.

Quand des ribosomes actifs sont isolés dans une fraction associée à des acides aminés marqués, ils sont trouvés sous la forme d'une unité constituée d'un ARNm et de plusieurs ribosomes. C'est le **polyribosome** ou **polysome**.

Chaque ribosome synthétise un seul polypeptide au cours de sa migration le long de la séquence qui porte le message. Ainsi, comme l'illustre la **Figure 5.5**, une série de ribosomes se déplace le long de la chaîne d'ARN, de l'extrémité 5' vers 3', chacun transportant une chaîne polypeptidique en voie de synthèse, de plus en plus longue.

Une configuration caractéristique de polysomes telle qu'elle apparaît en microscopie électronique est montrée **Figure 5.6**. La globine est synthétisée par des pentasomes. Chaque pentasome est constitué de 5 ribosomes reliés par un brin d'ARNm. Les ribosomes sont dispersés le long du messenger. Ceux qui sont situés à une extrémité commencent la synthèse protéique.; à l'autre extrémité

ils ont accompli la presque totalité de la synthèse de la chaîne polypeptidique.

La taille du polysome dépend de plusieurs facteurs. Chez les bactéries, elle peut être importante, puisqu'une dizaine de ribosomes peuvent traduire simultanément. La taille peut être due soit à la longueur de l'ARNm (qui peut coder pour plusieurs protéines), soit à l'efficacité avec laquelle les ribosomes traduisent l'ARNm. Comme les ribosomes se lient à l'ARNm bactérien avant même que sa transcription ne soit achevée, le polysome est sans doute également lié à l'ADN.

Chez les eucaryotes, l'ARNm doit sortir du noyau pour atteindre les ribosomes cytoplasmiques. Ils sont sans doute plus petits que les polysomes bactériens ; ici aussi, leur taille est fonction de la longueur de l'ARNm (il ne code que pour une seule protéine chez les eucaryotes) et de la fréquence particulière avec laquelle ils s'y unissent. En général, il est admis que moins de 10 ribosomes sont simultanément liés à l'ARNm.

Le nombre de ribosomes portés par une molécule d'ARNm, synthétisant une protéine particulière, n'est pas défini précisément, aussi bien chez les bactéries que chez les eucaryotes, mais est sujet à des variations statistiques reflétant deux paramètres : la taille de l'ARNm et son efficacité.

L'ARNm bactérien s'est révélé difficile à isoler car il ne représente qu'une très faible fraction de l'ARN total bactérien et il est de plus instable. Chez les bactéries, l'ARNm est synthétisé, traduit par les ribosomes et dégradé en une succession d'étapes rapides comme l'illustre la **Figure 5.7**. La durée de vie d'une molécule d'ARNm donnée peut être de quelques minutes ou même moins.

Chez les eucaryotes, l'ARNm représente également une faible proportion de l'ARN cellulaire total (environ 3 % du total). Par contre, il est relativement stable, sa durée de vie dans une cellule pouvant atteindre plusieurs heures.

Le **Tableau 5.1** donne un ordre d'idée de l'importance des différents éléments affectés à la synthèse protéique dans la bactérie intacte. Les quelque 20 000 ribosomes forment une proportion non négligeable de la masse cellulaire. Il y a dix fois moins de ribosomes que de molécules d'ARNt ; celles-ci sont presque toutes présentes sous forme d'aminoacyl-ARNt, c'est-à-dire prêtes à intervenir à tout instant dans la synthèse protéique. Le nombre de molécules d'ARNm est difficile à estimer du fait de leur instabilité mais on peut avancer un chiffre variant de 2 000 à 3 000 compte tenu des divers états de synthèse et de dégradation.

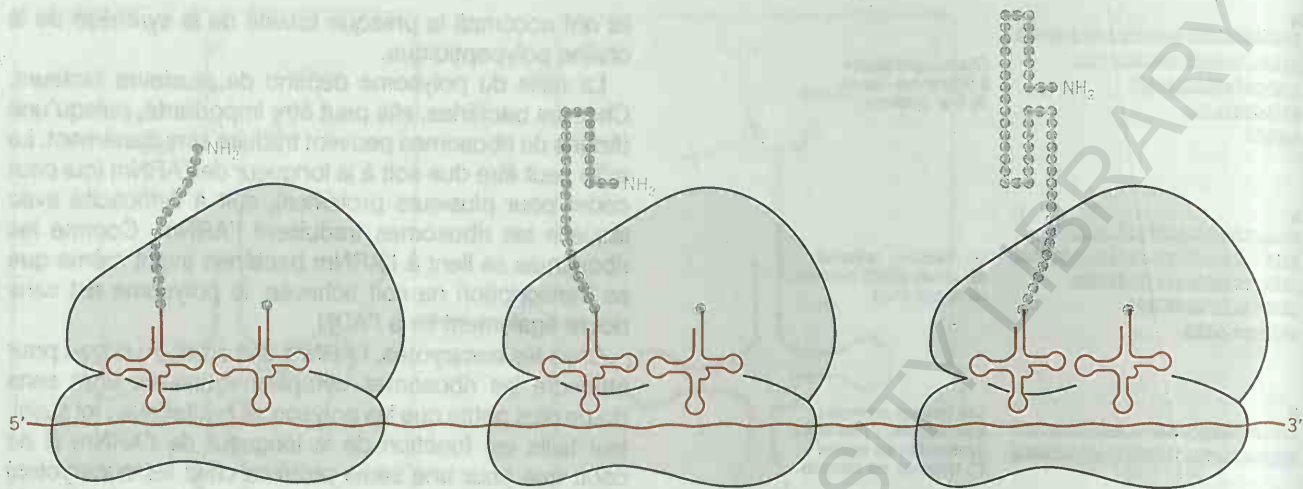


Figure 5.5

Un polyribosome est constitué d'un ARNm en cours de traduction simultanée par plusieurs ribosomes migrant dans la direction 5' vers 3'.

Chaque ribosome a deux molécules d'ARNt, la première porte le dernier acide aminé ajouté à la chaîne (à l'extrémité de la chaîne polypeptidique synthétisée jusque-là), la seconde transportant l'acide aminé ultérieur à ajouter.

La structure du ribosome protège de l'environnement environ les 30-35 derniers acides aminés ajoutés à la chaîne polypeptidique naissante. La partie précédente du polypeptide fait sans doute saillie et peut, en commençant à se replier, adopter sa conformation propre. (Donc, les protéines peuvent, avant même que leur synthèse ne soit achevée, partiellement adopter leur conformation mature).

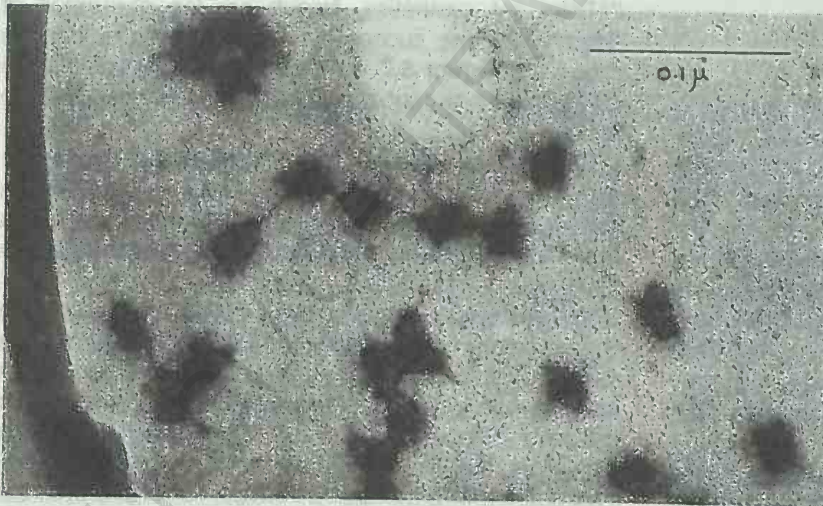


Figure 5.6

La synthèse protéique s'effectue toujours dans les polysomes.

La photographie en microscopie électronique montre des pentasomes synthétisant de la globine. Les ribosomes sont (en gros) des particules sphériques d'environ 7 nm (70Å) de diamètre, reliées par un fil d'ARNm. Selon les conditions, les polysomes peuvent apparaître compacts ou allongés.

Photographie aimablement fournie par Alex Rich.

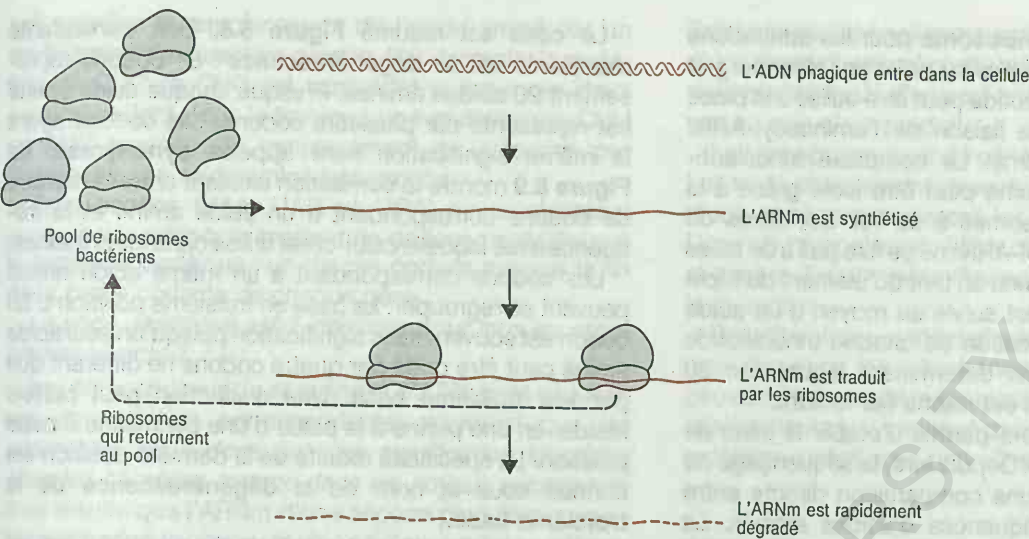


Figure 5.7
L'existence de l'ARN messenger dans les bactéries est éphémère. Un ARNm est synthétisé par transcription d'ADN puis est traduit par les ribosomes constitués à partir d'un pool de sous-unités libres. L'ARNm est rapidement dégradé ; les ribosomes réintègrent le pool et sont disponibles pour la traduction des ARNm ultérieurs.

LA SIGNIFICATION DU CODE GÉNÉTIQUE

Il y a 64 combinaisons de séquences de trinucléotides et seulement 20 acides aminés. Quand le code et sa lecture par codon ont été découverts, il était encore

Tableau 5.1
E. coli considéré en termes de constituants de la bactérie.

Constituant	Proportion de la masse cellulaire	Nombre/cellule
Paroi Membrane	10 % 10 %	
ADN	2 %	1
ARNm	2 %	2 500 ?
ARNt	3 %	160 000
ARNr	21 %	20 000
Protéine ribosomale	9 %	
Protéine soluble	40 %	10 ⁶
Petites molécules	3 %	5,5 × 10 ⁶

impossible d'attribuer un acide aminé aux codons. Peu après, deux méthodes ont permis de connaître la signification de chaque codon. Elles utilisaient des systèmes *in vitro*, composés des éléments nécessaires à la synthèse protéique.

Les systèmes de synthèse protéique peuvent être préparés à partir d'extraits cellulaires. La méthode initiale consistait principalement à casser les cellules d'*E. coli* et à centrifuger le mélange afin d'éliminer les parois bactériennes et les fragments membranaires. Si une source d'énergie adaptée et des précurseurs sont fournis au surnageant, celui-ci peut traduire tout ARNm qui lui est ajouté. (Des techniques identiques ont été utilisées depuis avec d'autres organismes et il existe maintenant plusieurs systèmes élaborés à haut degré de pureté.)

Le système *E. coli* a été utilisé pour traduire des polynucéotides synthétiques. En 1961, Nirenberg démontra le premier qu'avec un tel système, l'acide polyuridylique (poly U) agissait comme un ARNm dirigeant la synthèse de polyphénylalanine à partir de phénylalanine. Ce résultat montre que UUU doit être un codon pour la phénylalanine. Par la suite, Khorana a utilisé plusieurs polynucéotides synthétiques constitués de séquences connues de différentes bases ; ils ont permis d'attribuer une signification à presque la moitié des 64 codons.

L'essai de liaison au ribosome pour les attributions de sens aux codons a été mis au point par Nirenberg et Leder en 1964. Un trinucéotide peut être utilisé à la place d'un codon, entraînant la liaison de l'aminoacyl-ARNt correspondant au ribosome. Le complexe trinucéotide.aminoacyl-ARNt.ribosome peut être isolé grâce à la capacité qu'ont les ribosomes à se lier aux filtres de nitrocellulose. L'aminoacyl-ARNt ne se fixe pas à de telles membranes mais il est retenu en tant qu'élément du triple complexe. Sa rétention est suivie au moyen d'un acide aminé marqué. La signification de chaque trinucéotide peut être alors établie en déterminant lequel des 20 aminoacyl-ARNt marqués est retenu sur le filtre.

Ces deux techniques ont permis d'établir le sens de 61 des 64 codons *in vitro*. Depuis lors, le séquençage de l'ADN a rendu possible une comparaison directe entre les nucléotides et les séquences d'acides aminés. La séquence du brin codant de l'ADN, lue dans le sens 5' vers 3', est constituée de triplets qui correspondent à la séquence d'acides aminés de la protéine, lue de l'extrémité N-terminal à l'extrémité C-terminal. Le code génétique a été entièrement confirmé par ce type d'analyse.

Figure 5.8

Tous les codons triplets ont une signification.

Soixante et un codons représentent les acides aminés. Tous les acides aminés exceptés le tryptophane et la méthionine sont représentés par plus d'un codon. Les codons synonymes forment des groupes dans lesquels habituellement la base en troisième position possède le moins de signification. Trois codons provoquent la terminaison (TERM). L'ordre des bases dans un codon s'écrit de la même façon que pour les autres séquences, dans la direction 5' vers le 3'.

		DEUXIÈME BASE			
		U	C	A	G
PREMIÈRE BASE	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } TERM UAG }	UGU } Cys UGC } UGA } TERM UGG } Trp
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }
	A	AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }

Le code est résumé Figure 5.8. Une particularité marquante est sa **dégénérescence** : 61 codons représentent 20 acides aminés. Presque chaque acide aminé est représenté par plusieurs codons. Les codons ayant la même signification sont appelés **synonymes**. La Figure 5.9 montre la corrélation existant entre le nombre de codons correspondant à un acide aminé et la fréquence avec laquelle celui-ci est utilisé dans les protéines.

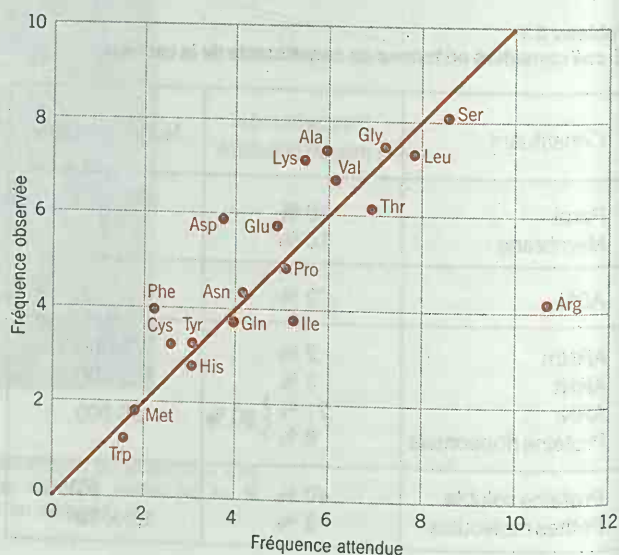
Les codons correspondant à un même acide aminé peuvent se regrouper. La base en troisième position d'un codon est souvent sans signification puisqu'un seul acide aminé peut être codé par quatre codons ne différant que par leur troisième base. Une distinction peut parfois résider en une purine à la place d'une pyrimidine à cette position. La spécificité réduite de la dernière position est connue sous le nom de la **dégénérescence de la troisième base**.

Cette particularité associée à la tendance qu'ont les acides aminés proches à être représentés par des codons voisins minimise les effets des mutations. Elle augmente en effet les chances que la mutation d'une base au hasard soit ne provoque pas la substitution d'un acide aminé,

Figure 5.9

Le nombre de codons pour chaque acide aminé est corrélé à sa fréquence d'utilisation dans les protéines.

L'arginine est une exception, car dans l'ADN eucaryote, la présence du doublet CG est réduite. Les quatre codons arginine qui commencent par ce doublet sont cependant présents dans l'ADN, mais à une fréquence inférieure à celle attendue du fait de leur composition en bases per se.



soit entraîne le remplacement de l'acide aminé par un acide aminé de caractère proche. Par exemple, la mutation de CUC en CUG est sans effet puisque ces deux codons codent pour la leucine ; une mutation de CUU en AUU provoque le remplacement de la leucine par l'isoleucine, un acide aminé très proche.

Trois codons (UAA, UAG et UGA) ne représentent aucun acide aminé. Ils mettent fin de façon spécifique à la synthèse protéique ; un de ces codons marque la fin de la partie codante de chaque gène.

Le code génétique est-il identique pour tous les organismes vivants ?

Les comparaisons de séquences d'ADN et de séquences de protéines correspondantes montrent que *les ensembles de codons ont une signification identique chez les bactéries comme dans les noyaux eucaryotes*. Il en résulte que l'ARNm d'une espèce peut être correctement traduit *in vitro* ou *in vivo* par l'appareil de synthèse protéique d'une autre espèce. Donc, les codons de l'ARNm d'une espèce ont la même signification pour les ribosomes et les ARNt d'autres espèces.

Du fait de la très grande variation de la composition en bases des différents génomes contrastant avec la stabilité générale du contenu en acides aminés des protéines, les différentes espèces utilisent les divers codons synonymes de chaque acide aminé à un degré variable et caractéristique de l'espèce. En fait, par sa dégénérescence, le code peut ainsi être préservé dans de telles circonstances.

L'universalité du code plaide en faveur de son établissement très tôt au cours de l'évolution. A l'origine une relation stéréochimique a pu exister entre les acides aminés et leurs codons respectifs. Puis, une fois le système de synthèse protéique créé, il a pu évoluer par une sélection de caractéristiques telles qu'une plus grande efficacité ou précision.

Il se peut que la forme primitive du code n'ait utilisé qu'un petit nombre de codons qui, comparativement, ne représentaient que peu d'acides aminés, avec peut-être un seul codon par groupe d'acides aminés. Une plus grande précision dans la signification des codons et de nouveaux acides aminés ont pu apparaître plus tardivement. Une possibilité, au début, était que seules deux bases sur trois étaient utilisées ; la discrimination au niveau de la troisième base serait apparue ultérieurement.

Le code actuel a pu être « gelé » à un moment donné, car l'ensemble devenait d'une complexité telle que tout changement dans le sens d'un codon aurait rompu les protéines existantes par substitution d'acides aminés.

Son universalité implique que ceci s'est produit au cours d'une étape d'une précocité telle que tous les organismes vivants sont issus d'un seul ensemble de cellules primitives où cela s'est produit.

Les exceptions au code génétique universel sont rares. Les seuls changements trouvés dans le génome principal d'une espèce concernent les codons de terminaison. Dans un mycoplasme, UGA code pour le tryptophane ; et dans le *Tetrahymena thermophila* cilié, UAA code pour la glutamine.

Des altérations systématiques du code ne sont apparues que dans les mitochondries. De tels changements peuvent être possibles car la mitochondrie est un système relativement clos, impliqué dans la synthèse de quelques protéines spécifiques (voir Chapitre 6).

LES SITES D'ACTION DU RIBOSOME

La synthèse protéique implique une chaîne de montage sur laquelle les ribosomes avancent inexorablement le long du messager, apportant les aminoacyls-ARNt qui fournissent les vrais modules de construction de la protéine. En lui-même, le ribosome est un petit atelier migrateur, à l'intérieur duquel un ensemble compact de protéines et d'ARNr forme plusieurs centres actifs pouvant entreprendre des activités catalytiques variées. Différentes batteries de facteurs annexes assistent le ribosome au cours des trois étapes de la synthèse protéique : l'initiation, l'élongation et la terminaison. L'énergie nécessaire à la synthèse protéique est apportée par l'hydrolyse du GTP.

L'initiation englobe les réactions qui précèdent la formation de la liaison peptidique entre les deux premiers acides aminés de la protéine. Elle nécessite la liaison du ribosome à l'ARNm avec formation d'un complexe d'initiation qui contient le premier aminoacyl-ARNt. Dans la synthèse protéique, cette étape est relativement lente.

L'élongation comprend toutes les réactions à partir de la formation de la première liaison peptidique jusqu'à l'addition du dernier acide aminé. Un par un, les acides aminés viennent s'ajouter à la chaîne protéique. L'addition d'un acide aminé est l'étape la plus rapide de la synthèse protéique.

La terminaison concerne les étapes nécessaires à la libération de la chaîne polypeptidique à complétion ; simultanément, le ribosome se dissocie de l'ARNm. La terminaison est lente comparée au temps d'addition d'un acide aminé à la chaîne.

Globalement, la synthèse protéique est un procédé rapide (bien que fortement dépendant de la température). Chez les bactéries, à 37 °C, la chaîne polypeptique s'accroît d'une quinzaine d'acides aminés toutes les secondes. Il suffit donc de 20 secondes pour synthétiser une protéine moyenne de 300 acides aminés. Chez les eucaryotes, la vitesse de synthèse est plus lente. Dans les globules rouges à 37 °C, l'élongation s'effectue classiquement au rythme de deux acides aminés par seconde.

La plupart des expériences qui permettent de déterminer les étapes de la synthèse protéique ont été réalisées *in vitro*, avec des systèmes comprenant les ribosomes, les aminoacyl-ARNt, les facteurs accessoires et une source d'énergie. Dans de tels systèmes, la vitesse de synthèse protéique peut être un peu plus lente qu'*in vivo*.

Les ribosomes sont généralement caractérisés par leur vitesse (approximative) de sédimentation. (L'unité de mesure de la vitesse de sédimentation est le Svedberg, en abrégé **S**. Plus la particule est lourde, plus la vitesse de sédimentation est élevée et plus la valeur de **S** est grande. La vitesse est également fonction de la forme, des particules plus compactes sédimenteront plus rapidement.) Les ribosomes bactériens sédimenteront habituellement à 70S. Les ribosomes cytoplasmiques des cellules eucaryotes supérieures, de plus grande taille, sédimenteront en général à 80S.

Un ribosome est constitué de deux sous-unités qui se dissocient *in vitro* quand la concentration en ions Mg^{2+} diminue. Les ribosomes (70S) bactériens ont deux sous-unités qui séimentent vers 50S et 30S. Les sous-unités des ribosomes cytoplasmiques eucaryotes (80S) séimentent aux environs de 60S et 40S. Dans chaque cas le rapport de taille entre les deux sous-unités est de l'ordre de deux. Les deux sous-unités agissent ensemble en tant que partie intégrante du ribosome mais chacune effectue des réactions distinctes dans la synthèse protéique.

A tout moment, l'ARNm est associé à la petite sous-unité par une trentaine de bases, mais seules deux molécules d'ARNt sont acceptées simultanément par le ribosome. Ainsi la synthèse protéique implique des réactions qui n'interviennent que sur deux des dix codons recouverts approximativement par le ribosome.

La Figure 5.10 montre que chaque ARNt est situé dans un site particulier. Les caractéristiques des deux sites sont différentes.

Le site A (ou site d'entrée) est le seul accès possible pour un aminoacyl-ARNt qui arrive. Avant l'entrée de l'aminoacyl-ARNt, le site expose le codon représentant l'acide aminé suivant qui doit être ajouté à la chaîne.

Le codon qui représente le *dernier* acide aminé devant être ajouté est situé au **site P** (ou **site donneur**). Ce site est occupé par un **peptidyl-ARNt**, c'est-à-dire l'ARNt chargé de la chaîne polypeptidique synthétisée en amont (voir *étape 1*).

Quand les deux sites sont occupés, une liaison peptidique s'établit par le transfert du polypeptide porté par le peptidyl-ARNt, vers l'acide aminé porté par l'aminoacyl-ARNt. Ce transfert s'effectue dans la grande sous-unité du ribosome.

L'extrémité de l'ARNt chargée de l'acide aminé est située dans la grande sous-unité tandis que l'anticodon, à l'autre extrémité, s'attache à l'ARNm lié à la petite sous-unité. Ainsi, les sites P et A se déplacent ensemble à travers les deux sous-unités ribosomales comme l'illustre la Figure 5.10.

Le transfert d'un polypeptide fait apparaître un ribosome dans lequel l'**ARNt désacylé**, dépourvu d'acide aminé, est au site P, tandis qu'un nouveau peptidyl-ARNt est créé au site A (voir *étape 2*). La chaîne peptidique de ce peptidyl-ARNt s'est allongée d'un acide aminé par rapport au peptidyl-ARNt qui était lié au site P dans l'*étape 1*.

Le ribosome se déplace ensuite d'un triplet sur le messager. Ceci chasse l'ARNt désacylé et permet au peptidyl-ARNt de se lier au site P (voir *étape 3*). Le nouveau codon à traduire est maintenant au site A, prêt à accueillir un nouvel aminoacyl-ARNt quand le cycle recommencera.

L'INITIATION NÉCESSITE LES SOUS-UNITÉS 30S ET DES FACTEURS ACCESSOIRES

En migrant le long de l'ARNm, afin d'allonger la chaîne polypeptidique, un ribosome catalyse la synthèse protéique. Mais le ribosome ne peut se lier directement à un ARNm pour initier la synthèse protéique. L'initiation met en évidence l'importance de chaque sous-unité du ribosome car elle requiert une petite sous-unité individuelle.

Quel est l'élément qui induit le début de la traduction d'une séquence ? Chaque ARNm possède un **site de liaison au ribosome**, courte séquence de bases qui précède la vraie région codante (voir Chapitre 8). Le contact initial avec l'ARNm intervient quand une petite sous-unité se fixe au site de liaison pour former un **complexe d'initiation**. En fait, cette région pourrait s'appeler site de liaison à la sous-unité 30S. Puis la grande sous-unité se joint au complexe pour constituer un ribosome complet.

Bien que la sous-unité 30S soit impliquée dans l'initia-

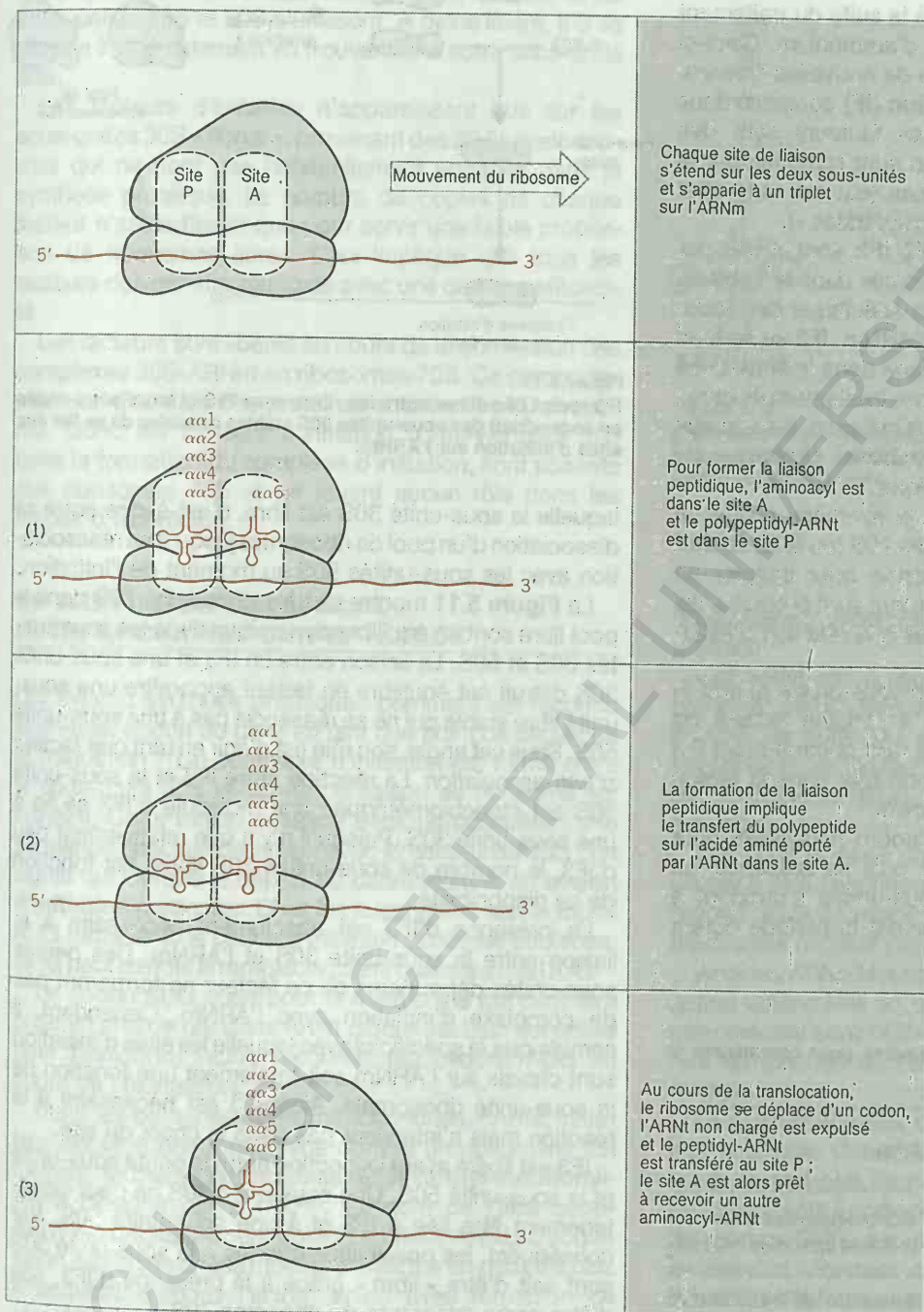


Figure 5.10
Le ribosome a deux sites de liaison pour l'ARNt.

tion, elle ne peut pas par elle-même entreprendre les réactions de liaison entre l'ARNm et l'ARNt initiateur. L'inefficacité des sous-unités ribosomales dans la réaction d'initiation a été découverte à la suite du traitement des sous-unités 30S au chlorure d'ammonium. Celui-ci les empêchait d'initier la synthèse de nouvelles chaînes. Ceci est dû aux **facteurs d'initiation (IF)** contenus dans les sous-unités 30S actives. Ces facteurs sont des protéines, dont les liaisons faibles sont rompues par le traitement au chlorure d'ammonium (qui est sans effet sur les protéines ribosomales « structurales »).

Trois facteurs d'initiation, **IF1**, **IF2**, **IF3**, sont utilisés par les bactéries. Leurs rôles sont résumés dans le **Tableau 5.2**. IF3 est nécessaire à la liaison spécifique des sous-unités 30S au site d'initiation de l'ARNm. IF2 se lie à un ARNt initiateur spécifique et le place dans la sous-unité 30S. Le rôle de IF1 n'est pas encore totalement élucidé ; en tant qu'élément du complexe d'initiation, il se lie aux sous-unités 30S et pourrait les stabiliser plutôt que de reconnaître un constituant spécifique.

Les ribosomes engagés dans la synthèse protéique existent sous la forme de particules 70S (ou 80S) et sont libérés de l'ARNm sous cette forme pour intégrer un groupe de ribosomes libres. Comment sont produites les sous-unités libres 30S nécessaires à la réaction d'initiation ?

Les ribosomes se dissocient en sous-unités quand ils se lient à une protéine appelée à l'origine **facteur de dissociation**, puis identifiée plus tard comme étant IF3 (découvert auparavant grâce à son rôle dans la liaison entre les sous-unités 30S et l'ARNm).

IF3 a deux fonctions : il est d'abord nécessaire pour engendrer les sous-unités 30S ; puis sa présence est indispensable pour que ces sous-unités s'unissent à l'ARNm. IF3 contrôle essentiellement la période durant

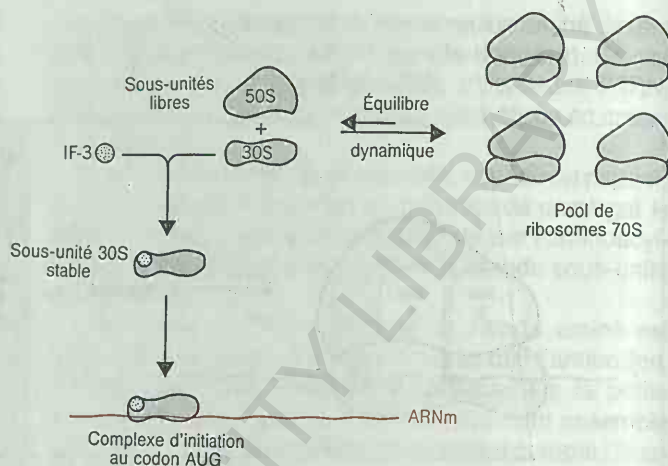


Figure 5.11

IF3 rompt l'équilibre entre les ribosomes 70S et leurs sous-unités 30S en engendrant des sous-unités 30S stables capables de se lier aux sites d'initiation sur l'ARNm.

laquelle la sous-unité 30S est libre, c'est-à-dire entre sa dissociation d'un pool de ribosomes jusqu'à sa réassociation avec les sous-unités 50S au moment de l'initiation.

La **Figure 5.11** montre que les ribosomes 70S dans le pool libre sont en équilibre dynamique avec les sous-unités 30S et 50S. La liaison entre un IF3 et une sous-unité 30S détruit cet équilibre en faisant apparaître une sous-unité libre stable qui ne se réassocie pas à une sous-unité 50S. Sous cet angle, son rôle est d'agir en tant que *facteur d'anti-association*. La réaction entre IF3 et la sous-unité 30S est stoechiométrique : une molécule d'IF3 se lie à une sous-unité 30S. Puisqu'il n'y a que relativement peu d'IF3, le nombre de sous-unités 30S libres est fonction de sa disponibilité.

La présence d'IF3 est absolument nécessaire à la liaison entre la sous-unité 30S et l'ARNm. Des petites sous-unités dépourvues de ce facteur ne formeront pas de complexe d'initiation avec l'ARNm. Cependant, il semble que la *spécificité* avec laquelle les sites d'initiation sont choisis sur l'ARNm soit totalement une fonction de la sous-unité ribosomale. Ainsi IF3 est nécessaire à la réaction mais n'intervient pas dans le choix du site.

IF3 est libéré avant la jonction entre la petite sous-unité et la sous-unité 50S. Une sous-unité 30S ne peut simultanément être liée à IF3 et à une sous-unité 50S. Par conséquent, les possibilités offertes à la sous-unité 30S sont soit d'être « libre » grâce à la présence d'IF3, soit d'être partie intégrante du ribosome 70S.

Au cours de son cycle de vie, la sous-unité 30S est soumise à ces alternatives. Quand un ribosome fait partie

Tableau 5.2

Trois facteurs d'initiation sont nécessaires pour commencer la synthèse protéique dans *E. coli*.

Facteur	Masse (daltons)	Facteurs /ribosome	Rôle
IF3	23 000	25 %	dissociation des sous-unités et liaison de l'ARNm
IF2	97 300	?	liaison de l'ARNt initiateur
IF1	9 000	15 %	recyclage ?

du pool libre, son équilibre dynamique avec les sous-unités permet à IF3 de remplacer la sous-unité 50S. Quand le complexe d'initiation est formé, IF3 est relâché, et les sous-unités 30S et 50S s'unissent. A peine libéré, IF3 se recycle immédiatement en trouvant une autre sous-unité 30S.

Les facteurs d'initiation n'apparaissent que sur les sous-unités 30S « libres », provenant des 20 % de ribosomes qui ne sont pas habituellement engagés dans la synthèse protéique. Le nombre de copies de chaque facteur n'est suffisant que pour servir une faible proportion de ribosomes libres. Ceci implique que tous les facteurs doivent être réutilisés avec une certaine efficacité.

Les facteurs sont libérés au cours de la conversion des complexes 30S-ARNm en ribosomes 70S. Ce comportement les distingue des protéines structurales du ribosome. Donc les facteurs d'initiation n'interviennent que dans la formation du complexe d'initiation, sont absents des ribosomes 70S et ne jouent aucun rôle dans les étapes d'élongation.

UN ARNt INITIATEUR SPÉCIAL INITIE LA CHAÎNE POLYPEPTIDIQUE

Qui initie la synthèse protéique : comment est reconnu le premier codon du gène en tant que point de départ de la traduction ? Un complexe d'initiation est formé entre la sous-unité 30S et le site de liaison au ribosome sur l'ARNm. Cette région possède deux particularités : une séquence spéciale de liaison avec la sous-unité 30S ; un signal qui indique le départ du cadre lecture, un **codon d'initiation** particulier. Chez les eucaryotes, seul le triplet AUG représente le codon d'initiation ; chez les bactéries, GUG peut parfois le remplacer, dans quelques cas UUG.

Le codon AUG code pour la méthionine qui peut être prise en charge par deux ARNt différents. L'un est utilisé pour l'initiation, l'autre reconnaît les codons AUG au cours de l'élongation.

Dans les bactéries et les mitochondries, l'ARNt initiateur est lié à une méthionine dont le groupe aminé est formylé, formant une molécule de **N-formyl-méthionyl-ARNt** désigné par $\text{ARNt}_{\text{Met}}^{\text{fMet}}$. Le nom de l'aminoacyl-ARNt est habituellement abrégé en fMet-ARNt. L'acquisition de l'acide aminé modifié s'effectue en deux étapes. Il acquiert d'abord l'acide aminé avec formation de Met-ARNt ; puis la réaction de formylation telle que l'illustre la **Figure 5.12** bloque le groupe aminé libre. L'acide aminé bloqué ne peut pas participer à l'élongation

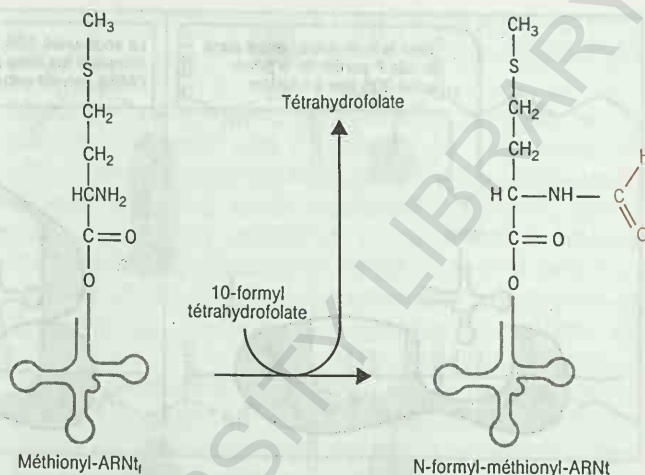


Figure 5.12

L'initiateur N-formyl-méthionyl-ARNt (fMet-ARNt) est produit par la formylation du méthionyl-ARNt, en utilisant comme cofacteur le formyl-tétrahydrofolate.

de la chaîne polypeptidique mais est capable d'initier une protéine.

Cet ARNt est *uniquement utilisé pour l'initiation* et reconnaît les codons AUG et GUG (parfois UUG). Les codons ne sont pas reconnus de façon équivalente : l'importance de l'initiation diminue de moitié quand AUG est remplacé par GUG, de moitié à nouveau quand UUG est utilisé.

Les codons AUG dont la localisation est interne ne peuvent être reconnus que par l' $\text{ARNt}_{\text{Met}}^{\text{fMet}}$. Cet ARNt répond *exclusivement aux codons AUG internes* ; sa méthionine ne peut pas être formylée.

Ainsi les ARNt-Met nécessaires à l'initiation et à l'élongation diffèrent-ils en deux points : les parties ARNt en elles-mêmes sont différentes et les acides aminés diffèrent en fonction de l'état de leur groupe NH_2 , libre ou formylé.

La signification des codons AUG et GUG dépend de leur **contexte**. Quand le codon AUG sert à l'initiation, il est lu en tant que formyl-méthionine ; à l'intérieur d'une région codante il code pour la méthionine. La signification du codon GUG est encore plus dépendante de sa localisation. *Premier codon*, il est lu via la réaction d'initiation comme une formyl-méthionine. A l'intérieur d'un gène, il est lu par l'ARNt-Val, un des ARNt courants, pour donner la valine selon les règles du code génétique (voir Figure 5.8).

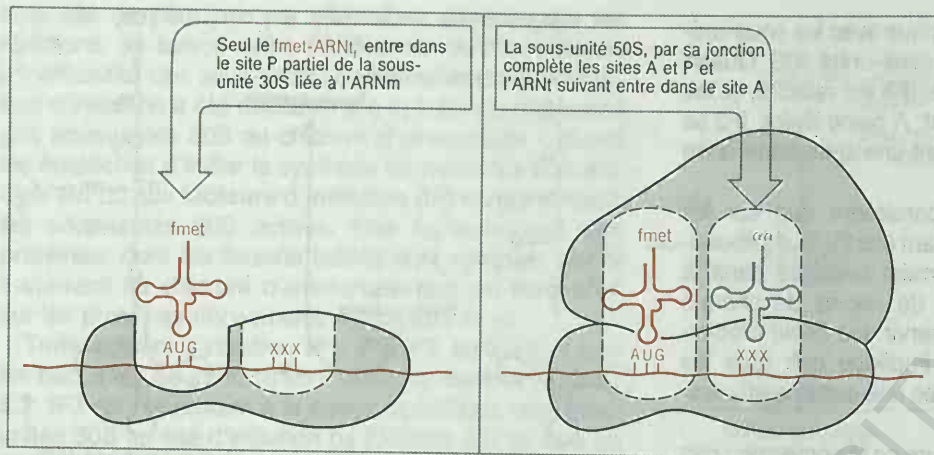


Figure 5.13

Seul le fMet-ARN_t est utilisé pour l'initiation par les sous-unités 30S ; seuls les autres aminoacyl-ARN_t sont utilisés pour l'élongation par les ribosomes 70S.

Comment est interprété le contexte des codons AUG et GUG ? La Figure 5.13 montre le rôle décisif du ribosome.

Dans un complexe d'initiation, la petite sous-unité est située sur l'ARN_m de telle sorte que le codon d'initiation est à l'intérieur du site P partiel qu'elle porte. Le seul amino-acyl ARN_t qui peut devenir membre du complexe d'initiation est l'initiateur, dont la propriété unique est de pouvoir pénétrer directement dans le site P pour reconnaître son codon.

Lorsque la grande sous-unité se joint au complexe, le fMet-ARN_t initiateur occupe le site P maintenant complet ; le site A est prêt à recevoir l'aminoacyl-ARN_t complémentaire du second codon du gène. Entre l'initiateur et l'aminoacyl-ARN_t suivant, la première liaison peptidique se forme. L'ARN_t initiateur se comporte comme un analogue d'un peptidyl-ARN_t (analogue dans la mesure où il fait don d'une chaîne peptidyl constituée d'un seul acide aminé).

Ainsi, l'initiation prédomine-t-elle quand un codon AUG (ou GUG) est voisin d'un signal qui indique que la petite sous-unité doit se lier *de novo* à l'ARN_m, puisque seul l'ARN_t initiateur peut pénétrer dans le site P partiel. En conséquence, la lecture interne l'emporte quand les codons sont parcourus par un ribosome *en train* de traduire un ARN_m, puisque seuls les aminoacyl-ARN_t habituels peuvent entrer dans le site A (complet).

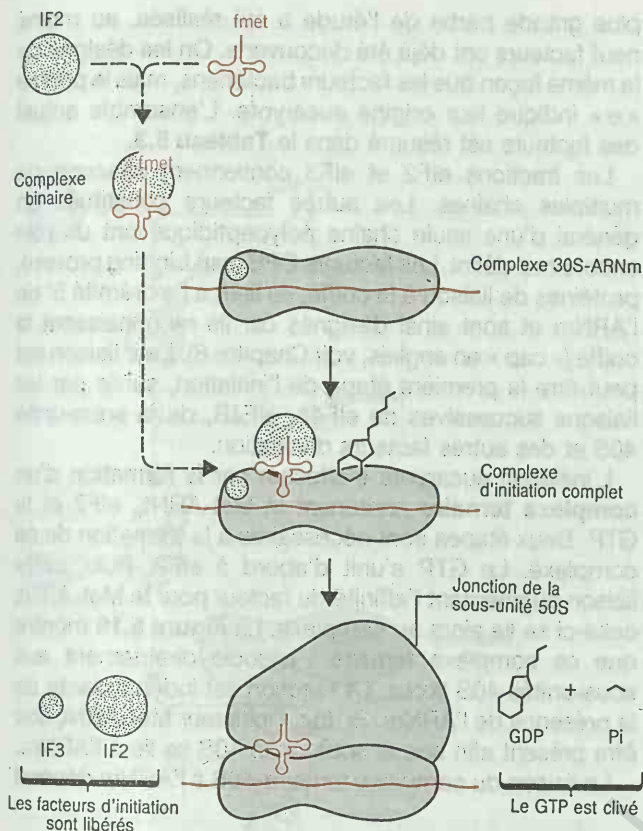
Des facteurs accessoires contrôlent la possibilité d'entrée des aminoacyl-ARN_t dans le ribosome. Le fMet-ARN_t initiateur est apporté au complexe 30S-ARN_m par IF2 comme l'illustre la Figure 5.14.

En formant un complexe spécifique avec le fMet-ARN_t, IF2 s'assure que c'est uniquement l'ARN_t initiateur et non un des aminoacyl-ARN_t usuels qui participe à la réaction d'initiation. En dehors de la présence essentielle d'un groupe aminé bloqué, les facteurs permettant la distinction entre l'ARN_t initiateur et les autres aminoacyl-ARN_t ne sont pas encore clairement établis.

Le complexe binaire IF2-fMet-ARN_t pourrait représenter la première étape de l'entrée de l'ARN_t initiateur au sein du complexe d'initiation. Puis le complexe 30S-ARN_m s'unit à IF2-fMet-ARN_t, introduisant l'ARN_t dans le site P partiel. L'ordre de ces réactions n'est pas prouvé ; IF2 peut se lier directement aux sous-unités 30S, de telle sorte qu'il se peut que fMet-ARN_t soit reconnu par IF2 déjà lié à la sous-unité 30S.

Quel que soit l'ordre de ces réactions, à cette étape, IF2 fait partie de la sous-unité 30S ; il doit ultérieurement remplir une autre fonction. On sait depuis longtemps qu'IF2 a une **activité GTPasique-ribosome dépendante** : il favorise l'hydrolyse du GTP en présence de ribosomes libérant l'énergie de cette liaison riche en énergie.

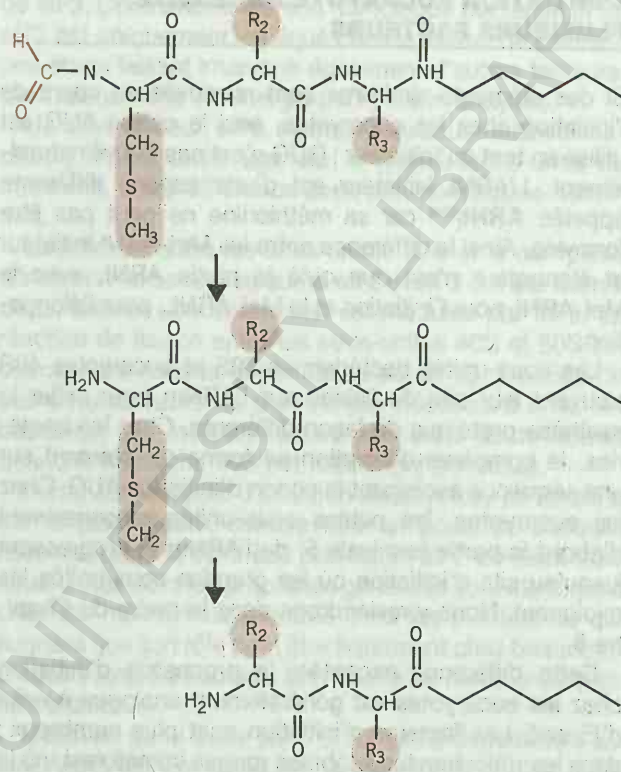
L'origine du GTP n'est pas évidente, mais il est probablement hydrolysé lorsque la jonction de la sous-unité

**Figure 5.14**

Les facteurs d'initiation bactériens sont nécessaires à la liaison entre les sous-unités 30S et l'ARNm et au fMet-ARNt, pour s'unir au complexe 30S-ARNm ; après la jonction avec la sous-unité 50S, les facteurs sont libérés et le GTP est clivé.

50S engendre un ribosome entier. IF2 n'est sans doute pas lui-même la GTPase mais active plutôt une protéine ribosomale qui possède cette activité. Le clivage du GTP implique peut-être un changement de conformation du ribosome afin que les deux sous-unités réunies soient converties en un ribosome 70S actif.

Après le clivage du GTP, les sites A et P du ribosome sont en position correcte. Le fMet-ARNt, est au site P ; le site A est prêt à recevoir l'aminoacyl-ARNt de façon normale. Dans ce modèle, l'hydrolyse du GTP a un rôle analogue à la rupture de la liaison riche en énergie qui permet le déplacement du ribosome (voir plus loin).

**Figure 5.15**

Les protéines naissantes commencent par la formyl-méthionine, mais le groupe formylé (et quelquefois la méthionine) est libéré au cours de la synthèse protéique.

Chez les bactéries et dans les mitochondries, le résidu formylé est clivé par une déformylase spécifique avec formation d'un N-terminal normal. Si la méthionine doit être l'acide aminé N-terminal de la protéine, c'est la seule étape nécessaire. Pour environ la moitié des protéines, la méthionine terminale est détachée par une aminopeptidase, créant le nouvel R₂ terminal (à l'origine le second acide aminé incorporé dans la chaîne). Quand ces deux étapes sont nécessaires, elles s'effectuent séquentiellement.

Le rôle de IF1 n'est pas clair. Il pourrait s'agir d'un **facteur de recyclage** qui interviendrait en aidant à libérer IF2.

Puisque le même acide aminé sert à initier toutes les chaînes protéiques, il est quelquefois nécessaire de se débarrasser des résidus indésirables en N-terminal, comme l'illustre la **Figure 5.15**. Cette (ces) réaction(s) a (ont) lieu assez précocement, probablement quand la chaîne polypeptidique naissante n'est constituée que de 15 à 30 acides aminés.

L'INITIATION EUCARYOTE NÉCESSITE PLUSIEURS FACTEURS

Si des éléments similaires sont retrouvés au cours de l'initiation chez les eucaryotes, seul le codon AUG est utilisé en tant qu'initiateur ; GUG n'est pas trouvé naturellement. L'ARNt initiateur est d'une espèce différente appelée **ARNt_i^{Met}** car sa méthionine ne peut pas être formylée. Ainsi la différence entre les Met-ARNt initiateur et élongateur n'est due qu'à la partie ARNt, avec le Met-ARNt_i pour l'initiation et le Met-ARNt_m pour l'élongation.

Les sous-unités bactériennes 30S et eucaryotes 40S trouvent leur site de liaison sur l'ARNm pour initier la synthèse protéique de façon différente. Chez les bactéries, le complexe d'initiation se forme directement sur une séquence avoisinant le codon d'initiation AUG. Chez les eucaryotes, les petites sous-unités reconnaissent d'abord la partie terminale 5' de l'ARNm et progressent jusqu'au site d'initiation où les grandes sous-unités les rejoignent. Nous y reviendrons dans le cadre du Chapitre 8.

Cette différence exceptée, le processus d'initiation chez les eucaryotes est généralement analogue à celui d'*E. coli*. Les facteurs d'initiation sont plus nombreux ; dans les réticulocytes (globules rouges immatures), où la

plus grande partie de l'étude a été réalisée, au moins neuf facteurs ont déjà été découverts. On les désigne de la même façon que les facteurs bactériens, mais le préfixe « e » indique leur origine eucaryote. L'ensemble actuel des facteurs est résumé dans le **Tableau 5.3**.

Les fractions eIF2 et eIF3 contiennent chacune de multiples chaînes. Les autres facteurs constitués en général d'une seule chaîne polypeptidique ont un rôle encore mal défini. Les facteurs CPB (cap binding protein), protéines de liaison à la coiffe, se lient à l'extrémité 5' de l'ARNm et sont ainsi désignés car ils reconnaissent la coiffe (« cap » en anglais, voir Chapitre 8). Leur liaison est peut-être la première étape de l'initiation, suivie par les liaisons successives de eIF4A, eIF4B, de la sous-unité 40S et des autres facteurs d'initiation.

L'initiation eucaryote s'effectue par la formation d'un **complexe ternaire** contenant le Met-ARNt_i, eIF2 et le GTP. Deux étapes sont nécessaires à la formation de ce complexe. Le GTP s'unit d'abord à eIF2. Puis, cette liaison augmentant l'affinité du facteur pour le Met-ARNt_i, celui-ci se lie alors au complexe. La **Figure 5.16** montre que ce complexe ternaire s'associe directement aux sous-unités 40S libres. La réaction est indépendante de la présence de l'ARNm. En fait, l'initiateur Met-ARNt_i doit être présent afin que la sous-unité 40S se lie à l'ARNm.

La liaison du complexe ternaire-40S à l'ARNm dépend

Tableau 5.3

Les réticulocytes contiennent au moins neuf facteurs d'initiation.

Facteur	Structure	Rôle
eIF3 CBPI CBPII	> 500 000 multimère 24 000 monomère 200 000 monomère	liaison de l'ARNm liaison de l'ARNm en 5' et désenroulement "
eIF1 eIF4B eIF4A	15 000 monomère monomère monomère	aide à la liaison de l'ARNm " " et lie l'ATP
eIF6	23 000 monomère	prévient la jonction 40 S-60 S
eIF5 eIF4C	150 000 monomère monomère	libération d'eIF2 et eIF3 liaison de la sous-unité 60 S
eIF2	trimère	liaison du Met ARNt _i (voir Tableau 5.4)
eIF4D	monomère	inconnu

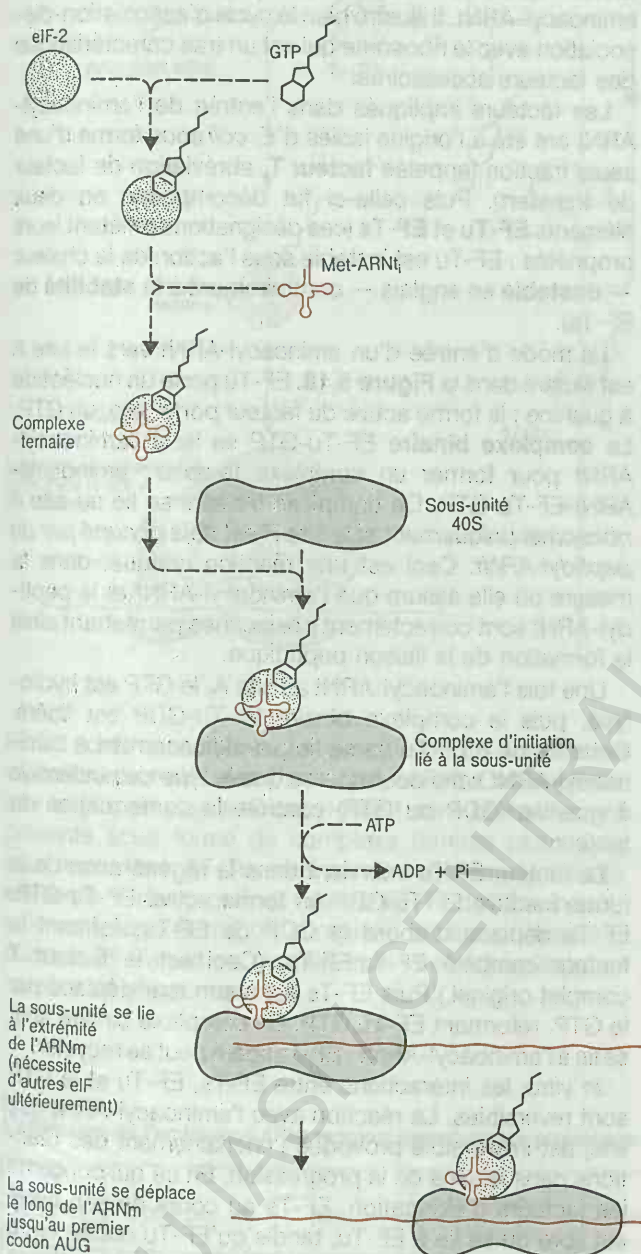


Figure 5.16
Dans l'initiation eucaryote, eIF2 forme un complexe ternaire avec le fMet-ARN_i. Ce complexe se lie aux sous-unités libres 40S qui s'unissent à l'extrémité 5' de l'ARNm et se déplacent vers le site d'initiation.

de eIF3. Contrairement à son homologue bactérien (IF3), eIF3 est *uniquement* impliqué dans la liaison de l'ARNm, une étape faisant intervenir également d'autres facteurs. Une liaison riche en énergie (ATP) est hydrolysée.

Un autre facteur, eIF6, maintient les sous-unités dissociées ; il peut aussi agir sur la grande sous-unité du ribosome, autre différence avec son homologue bactérien.

La jonction entre les sous-unités 60S et le complexe d'initiation ne se réalise que si eIF2 et eIF3 ont été libérés de ce dernier ; cette réaction est modulée par eIF5. La réaction de liaison entre les sous-unités 40S et 60S est directement sous la dépendance de eIF4C. L'action de ces deux facteurs ne s'avère pas nécessaire chez les bactéries. Les facteurs restant sont sans doute libérés quand le ribosome 80S est totalement constitué.

Parmi les facteurs d'initiation eucaryotes, eIF2, qui a été le plus étudié, semble être un point de contrôle essentiel dans la synthèse protéique. eIF2 contient trois sous-unités dont les propriétés générales sont identiques chez les mammifères comme chez les plantes, ce qui suggère que son rôle peut être équivalent chez beaucoup d'eucaryotes. Le **Tableau 5.4** résume les caractéristiques de ce facteur.

L'activité de la sous-unité γ qui se lie directement au Met-ARN_i est contrôlée par la sous-unité α . eIF2 n'est actif que si la sous-unité α porte un GTP. Au cours de l'initiation, le GTP est hydrolysé en GDP ; le GTP doit être régénéré afin que le facteur puisse agir à nouveau. Un autre facteur, appelé eIF2B, est nécessaire à cette réaction résumée dans la **Figure 5.17**.

La phosphorylation de la sous-unité α intervient dans différentes conditions. Elle bloque l'action de eIF2. En fait, la phosphorylation de 20 à 30 % seulement de eIF2 suffit à stopper l'initiation de la synthèse protéique,

Tableau 5.4
eIF2 est constitué de trois sous-unités.

Sous-unité	Masse (daltons)	Rôle dans l'initiation
α	35 000	lie le GTP contrôlée par la phosphorylation
β	38 000	peut être un facteur de recyclage
γ	55 000	lie le Met-ARN _i

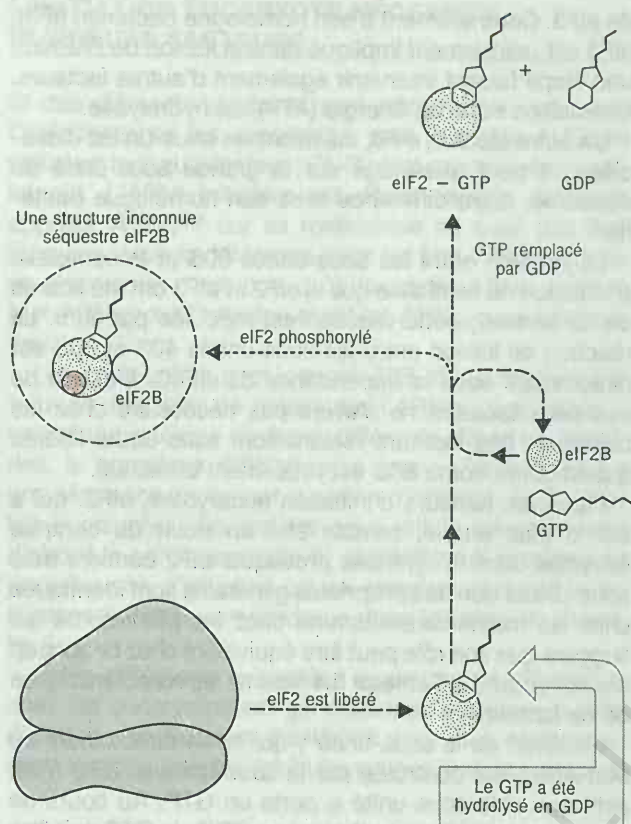


Figure 5.17
eIF2 est libéré à la fin de la réaction d'initiation eucaryote sous forme d'un complexe binaire contenant le GDP ; le facteur eIF2B est nécessaire pour remplacer le GDP par le GTP afin que eIF2 puisse être utilisé à nouveau.

probablement du fait que eIF2 phosphorylé se lie à toutes les molécules de eIF2B (10-20 % de la concentration de eIF2), les empêchant de participer à la régénération du GTP.

LE FACTEUR D'ÉLONGATION T APPORTE UN AMINOACYL-ARNt DANS LE SITE A

Quand le ribosome complet est constitué au codon d'initiation, la scène est prête pour un cycle au cours duquel l'aminocyl-ARNt entrera au site A du ribosome dont le site P est occupé par un peptidyl-ARNt. *Excepté l'initiateur, tout aminocyl-ARNt peut s'intégrer au site A.* Son entrée est modulée par un **facteur d'élongation (EF)**.

Comme sa contrepartie dans l'initiation (IF2), ce facteur n'est associé au ribosome que lorsqu'il favorise l'entrée de l'aminocyl-ARNt. Quand ce dernier est en place, le facteur quitte le ribosome pour agir avec un nouvel aminocyl-ARNt. Il illustre bien le cycle d'association-dissociation avec le ribosome qui est un trait caractéristique des facteurs accessoires.

Les facteurs impliqués dans l'entrée de l'aminocyl-ARNt ont été à l'origine isolés d'*E. coli* sous forme d'une seule fraction (appelée **facteur T**, abréviation de facteur de transfert). Puis celle-ci fut décomposée en deux éléments **EF-Tu** et **EF-Ts** (ces désignations reflètent leurs propriétés : EF-Tu est instable sous l'action de la chaleur — **unstable** en anglais — contrairement à la **stabilité** de EF-Ts).

Le mode d'entrée d'un aminocyl-ARNt vers le site A est illustré dans la **Figure 5.18**. EF-Tu porte un nucléotide à guanine ; la forme active du facteur porte, elle, un GTP. Le **complexe binaire** EF-Tu-GTP se lie à l'aminocyl-ARNt pour former un complexe ternaire : aminocyl-ARNt-EF-Tu-GTP. Ce complexe ternaire se lie au **site A ribosomal uniquement si le site P est déjà occupé par un peptidyl-ARNt**. Ceci est une réaction critique, dans la mesure où elle assure que l'aminocyl-ARNt et le peptidyl-ARNt sont correctement positionnés permettant ainsi la formation de la liaison peptidique.

Une fois l'aminocyl-ARNt au site A, le GTP est hydrolysé, puis le complexe binaire EF-Tu-GDP est libéré. Cette forme d'EF-Tu ne se lie pas efficacement à l'aminocyl-ARNt sans doute parce que le type de nucléotide à guanine (GDP ou GTP) contrôle la conformation du facteur.

Le facteur EF-Ts intervient dans la régénération de la forme inactive EF-Tu-GDP en forme active EF-Tu-GTP. EF-Ts déplace d'abord le GDP de EF-Tu, formant le facteur combiné EF-Tu.EF-Ts. (Ceci est le facteur T complet originel.) Puis EF-Ts est à son tour déplacé par le GTP, reformant EF-Tu.GTP. Ce complexe binaire actif se lie à l'aminocyl-ARNt ; EF-Ts libéré peut se recycler.

In vitro, les interactions entre EF-Ts, EF-Tu et le GTP sont réversibles. La réaction avec l'aminocyl-ARNt qui, elle, est irréversible provoque l'enchaînement des réactions dans le sens de la progression. En ce qui concerne les facteurs d'élongation, EF-Ts au cours de son cycle est libre ou se lie à EF-Tu, tandis qu'EF-Tu oscille entre une liaison au GDP ou au GTP dans les complexes binaires ou ternaires et l'association à EF-Ts au sein du facteur T.

Dans le **Tableau 5.5**, quelques propriétés de EF-Tu et

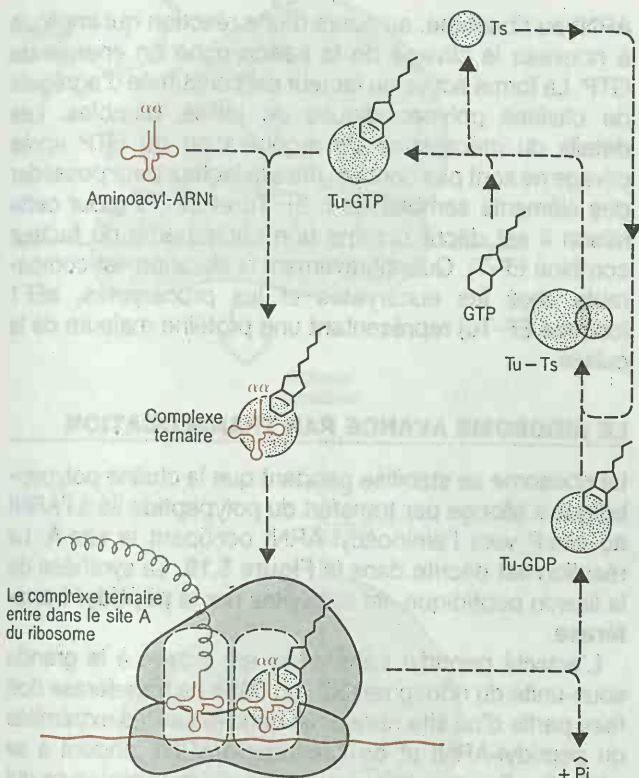


Figure 5.18
EF-Tu.GTP place l'aminocyl-ARNt sur le ribosome puis est libéré sous forme de EF-Tu.GDP, qui requiert EF-Ts comme intermédiaire dans le remplacement du GDP par le GTP.

Dans l'entrée au sein du complexe ternaire, la caractéristique structurale la plus importante de l'ARNt semble être le bras accepteur (la séquence double-brin se terminant par -ACCA-OH auquel l'acide aminé est lié). Que l'acide aminé soit lié au 2' ou 3' du noyau osidique est sans importance, mais au cours d'une étape qui précède la liaison au ribosome, le complexe ternaire stabilise la liaison ester de l'aminocyl-ARNt en position 2'.

Le seul aminocyl-ARNt qui ne peut être reconnu par EF-Tu.GTP est le fMet-ARNt, ce qui ne lui permet pas de répondre aux codons internes AUG ou GUG. A l'origine de cette impossibilité il y a peut-être la structure inhabituelle de son bras accepteur, qui diffère de celle des autres aminocyl-ARNt en ce sens qu'elle se termine par un groupe A.C non apparié au lieu d'une paire A.U. Le blocage du groupe NH₂ est lui aussi important. (Les caractéristiques exactes responsables du maintien de cette discrimination peuvent être spécifiques des bactéries, bien que le même principe s'applique aux eucaryotes).

EF-Ts sont rassemblées. La quantité de EF-Tu avoisine le nombre de molécules d'aminocyl-ARNt. Ceci implique que la plupart des aminocyl-ARNt sont probablement présents sous forme de complexe ternaire plutôt que sous forme libre. EF-Tu, codé par deux gènes, représente environ 5 % des protéines totales. EF-Tu doit être principalement sous forme de complexe binaire ou ternaire, puisque EF-Ts est en quantité moindre, équivalente au nombre de ribosomes.

Le rôle du GTP dans le complexe ternaire a été étudié en lui substituant un analogue qui ne peut être hydrolysé. Cet analogue est le **GMP-PCP** qui possède un pont méthylène à la place de l'oxygène qui lie les phosphates β et γ du GTP. En présence de GMP-PCP le complexe ternaire peut se former et lier l'aminocyl-ARNt au ribosome. Mais la liaison peptidique ne peut se réaliser. Ainsi la fixation de l'aminocyl-ARNt au site A requiert la

Tableau 5.5
L'utilisation de l'aminocyl-ARNt requiert deux facteurs protéiques.

Facteur	Codé par	Masse	Molécules /Cellule	Fonction	Inhibiteur
EF-Tu	<i>tufA, tufB</i>	43 225	70 000	lie aminocyl-ARNt et GTP	kirromycine
EF-Ts	<i>tsr</i>	74 000	10 000	lie EF-Tu en déplaçant GDP	

présence de GTP mais l'*hydrolyse* n'est nécessaire qu'ultérieurement.

Une grande partie de la compréhension des étapes de la synthèse protéique bactérienne a été obtenue par l'utilisation des antibiotiques qui inhibent cette synthèse à des niveaux particuliers. Les propriétés de certains inhibiteurs, les plus largement utilisés, sont résumées dans le **Tableau 5.6**.

La **kirromycine** inhibe l'action de EF-Tu. Lié à la kirromycine, EF-Tu peut encore unir un aminoacyl-ARNt au site A. Mais le complexe EF-Tu.GDP ne peut se libérer du ribosome. Sa présence permanente prévient la formation de la liaison peptidique entre le peptidyl-ARNt et l'aminoacyl-ARNt. Il en résulte que le ribosome « cale » sur l'ARNm entraînant un arrêt de la synthèse protéique.

L'action de la kirromycine montre que l'inhibition d'une étape de la synthèse protéique bloque l'étape suivante. Dans ce cas, EF-Tu.GDP doit être libéré pour que le ribosome puisse acquérir une conformation correcte qui favorise la liaison peptidique. Un principe identique régit les autres étapes de la synthèse protéique : une réaction doit être convenablement achevée avant que la suivante puisse se réaliser.

Chez les eucaryotes, dont le type est le système réticulocytaire, deux facteurs d'élongation existent, **eEF1** et **eEF2**, respectivement analogues des facteurs bactériens, EF-T et EF-G (voir plus loin).

Le rôle de eEF1 consiste à faire parvenir un aminoacyl-

ARNt au ribosome, au cours d'une réaction qui implique à nouveau le clivage de la liaison riche en énergie du GTP. La forme active du facteur est constituée d'agrégats de chaînes polypeptidiques de tailles variables. Les détails du mécanisme de régénération du GTP après clivage ne sont pas connus, mais le facteur peut posséder des éléments semblables à EF-Tu et EF-Ts (pour cette raison il est décrit comme une contrepartie du facteur combiné EF-T). Quantitativement la situation est comparable chez les eucaryotes et les procaryotes, eEF1 (comme EF-Tu) représentant une protéine majeure de la cellule.

LE RIBOSOME AVANCE PAR TRANSLOCATION

Le ribosome se stabilise pendant que la chaîne polypeptidique s'allonge par transfert du polypeptide lié à l'ARNt au site P vers l'aminoacyl-ARNt occupant le site A. La réaction est décrite dans la **Figure 5.19**. La synthèse de la liaison peptidique est catalysée par la **peptidyl transférase**.

L'activité peptidyl transférase est propre à la grande sous-unité du ribosome (50S ou 60S). La transférase doit faire partie d'un site ribosomal dans lequel les extrémités du peptidyl-ARNt et de l'aminoacyl-ARNt tendent à se rejoindre. Peu de choses sont encore connues en ce qui concerne les éléments réels qui possèdent cette activité transférase. On peut déduire des effets de l'érythromy-

Tableau 5.6

Les inhibiteurs de la synthèse protéique agissent sur une grande variété d'étapes.

Cible	Inhibiteur	Activité dans	Action ou nature des mutations résistantes
Initiation	kasugamycine streptomycine	bactérie bactérie	mutants sans groupements méthyle dans l'ARNr 16S mutations dans la protéine S12 de la sous-unité 30S
Élongation	kirromycine puromycine	bactérie bactérie	EF-Tu•GDP ne peut être libéré accepte le polypeptide à partir du peptidyl-ARNt
Peptidyl transférase	érythromycine chloramphénicol cycloheximide	bactérie et mitochondrie bactérie et mitochondrie cytoplasme eucaryote	mutations dans la modification de l'ARNr 23S mutations dans la séquence de l'ARNr 23S mutations dans les protéines 50S mutations dans la séquence de l'ARNr 23S inhibe la peptidyl transférase dans le 60S
Translocation	acide fusidique thiostreptone	bactérie bactérie	EF-G•GDP ne peuvent pas être libérés inhibe l'activité GTPase

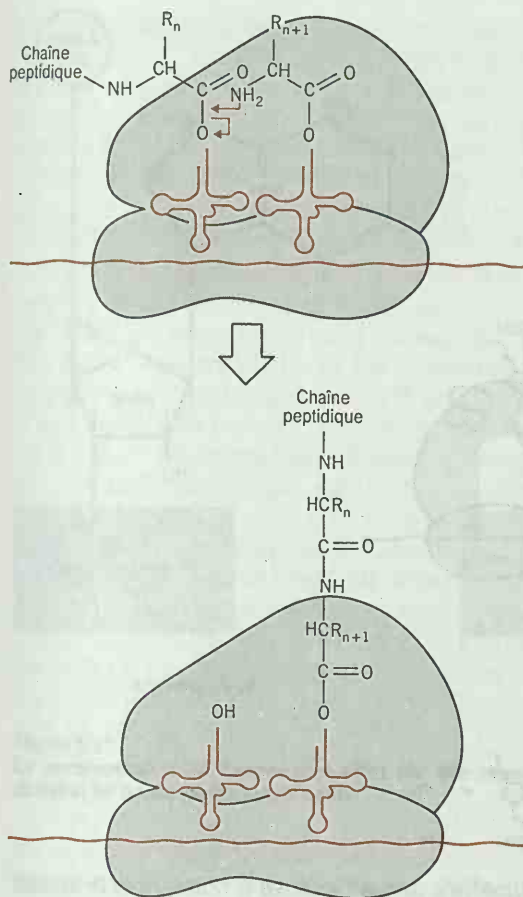


Figure 5.19

La formation de la liaison peptidique s'effectue par le transfert du polypeptide lié à l'ARNt dans le site P, à l'aminoacyl-ARNt dans le site A.

La peptidyl transférase préfère l'aminoacyl-3'-ARNt comme substrat. L'acide aminé est probablement transféré de sa position privilégiée en 2' dans le complexe ternaire avant que la liaison peptidique ne se forme.

Une fois la liaison peptidique établie, le ribosome possède un ARNt non chargé au site P et un peptidyl-ARNt au site A. Le cycle d'addition des acides aminés à la chaîne polypeptidique naissante s'achève par la **translocation**, comme l'illustre la Figure 5.20, phase au cours de laquelle le ribosome avance de trois nucléotides sur l'ARNm. La translocation simultanément chasse l'ARNt non chargé et place le peptidyl-ARNt dans le site P. Ainsi, le site A du ribosome, non occupé, est prêt à recevoir l'aminoacyl-ARNt qui correspond au codon suivant.

La nécessité d'un facteur supplémentaire au cours de la translocation fut très tôt mise en évidence grâce à la **puromycine**, un antibiotique dont la structure est assez proche de celle d'un acide aminé lié à l'adénosine terminale d'un ARNt. (En fait, ce sont ces expériences qui ont permis à l'origine de faire la distinction entre les sites A et P).

La Figure 5.21 montre que la puromycine a un N à la place du O qui unit l'acide aminé à l'ARNt. Le ribosome traite l'antibiotique comme un aminoacyl-ARNt entrant. Le polypeptide lié au peptidyl-ARNt est transféré sur le groupe NH_2 de la puromycine. Comme la liaison polypeptide-puromycine ne peut être clivée, le polypeptide ne peut pas être transféré vers l'aminoacyl-ARNt suivant. Il se sépare alors du ribosome sous forme de polypeptidyl-puromycine. L'effet létal de cet antibiotique résulte de l'arrêt prématuré de la synthèse protéique.

La puromycine qui accède au ribosome de la même façon que l'aminoacyl-ARNt ne réagit avec un peptidyl-ARNt que lorsque le site P est occupé et le site A libre (cas représenté dans la partie inférieure de la Figure 5.21). Quand le pont peptidique s'est formé mais que la translocation n'a pas encore eu lieu, la puromycine ne peut entrer car le site A reste occupé par le peptidyl-ARNt (partie supérieure de la Figure 5.21).

Dans ce dernier cas, la translocation est nécessaire pour rendre le ribosome sensible à l'action de la puromycine. *In vitro*, l'acquisition de cette sensibilité requiert la présence de GTP et d'un autre facteur d'élongation, **EF-G**. (L'appellation de cette protéine provient de son isolement en tant que facteur dépendant du GTP.) EF-G est un constituant majeur de la cellule; comme le Tableau 5.7 le résume, il est présent à raison d'une copie environ par ribosome.

EF-G favorise la translocation en se liant au ribosome; il est ensuite libéré quand le GTP est hydrolysé. La substitution du GTP par le GMP-PCP conserve la possibilité de liaison, ce qui indique que la *présence* d'un

cine et du chloramphénicol, résumés dans le Tableau 5.6, que l'ARNr 23s et les sous-unités protéiques 50S sont nécessaires.

Il reste encore à comprendre dans la synthèse protéique, comment deux molécules d'ARNt, sans aucun doute volumineuses, peuvent avoir les bases de leur anticodon appariées à des codons adjacents sur l'ARNm, tandis que leurs sites accepteurs sont suffisamment proches pour permettre la formation d'une liaison peptidique. Nous ne savons pas encore comment la géométrie des sites P et A contourne cet obstacle stérique présumé.

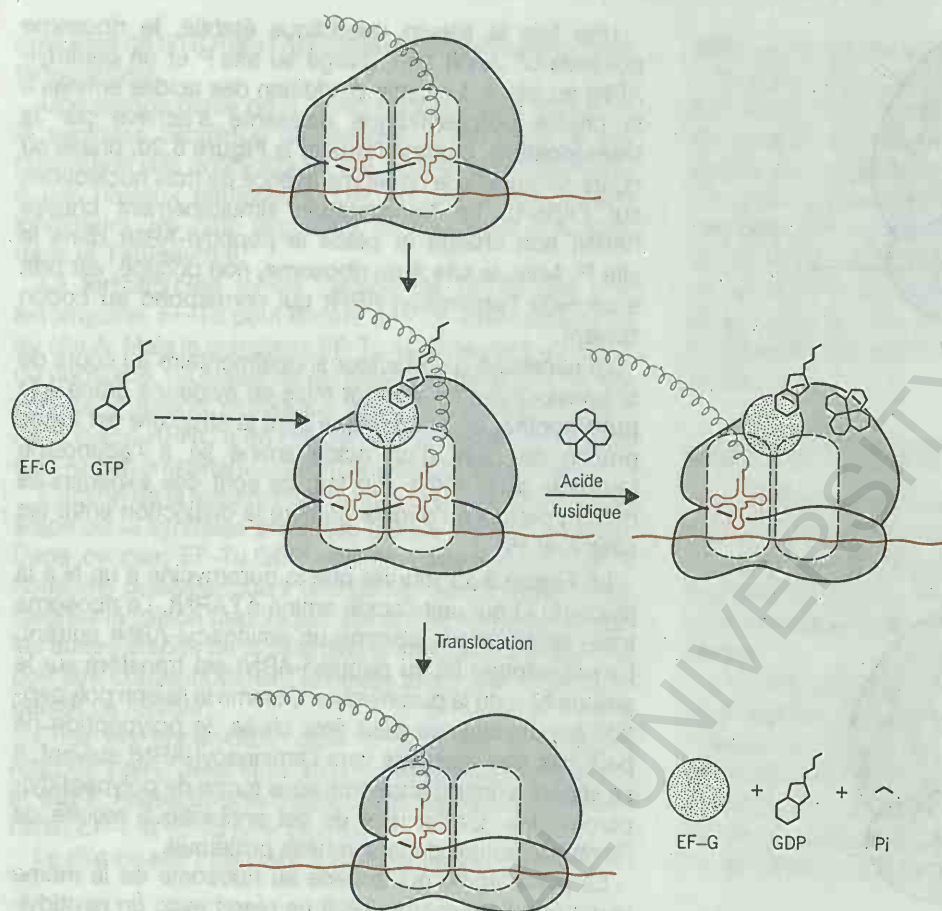


Figure 5.20

Une fois la liaison peptidique réalisée, la translocation nécessite EF-G et le GTP. Habituellement, EF-G et le GDP sont libérés, mais l'acide fusidique peut stabiliser le complexe en situation de post-translocation.

nucléotide à guanine est nécessaire à cette liaison mais son *hydrolyse* n'est pas utile tant qu'il n'y a pas translocation. On ne sait pas si le GTP nécessaire à la translocation est fourni au ribosome par EF-G ou d'une autre façon. L'hydrolyse du GTP n'est pas catalysée par le facteur mais est une fonction du ribosome.

Puisque les ribosomes ne peuvent être liés simultanément à EF-Tu et EF-G, la synthèse protéique suit un cycle au cours duquel les facteurs sont alternativement liés et relâchés des ribosomes. Ainsi, EF-Tu.GDP doit être libéré avant que EF-G ne se lie, ce dernier devant être libéré avant que l'aminocyl-ARNt.EF-Tu.GTP puisse se lier.

L'action de l'acide fusidique, antibiotique stéroïde, qui « bloque » le ribosome en post-translocation a permis de mettre en évidence l'utilité de la libération de EF-G. En présence d'acide fusidique, un cycle de translocation s'effectue : EF-G se lie au ribosome, le GTP est hydrolysé et le ribosome avance de trois nucléotides. Mais l'acide fusidique stabilisant le complexe ribosome.EF-G.GDP, EF-G et GDP restent associés au ribosome au lieu d'être libérés (voir Figure 5.20). Comme le ribosome ne peut plus se lier à un aminocyl-ARNt, aucun acide aminé ne peut venir allonger la chaîne peptidique.

Nous ne savons pas encore si la capacité de chaque

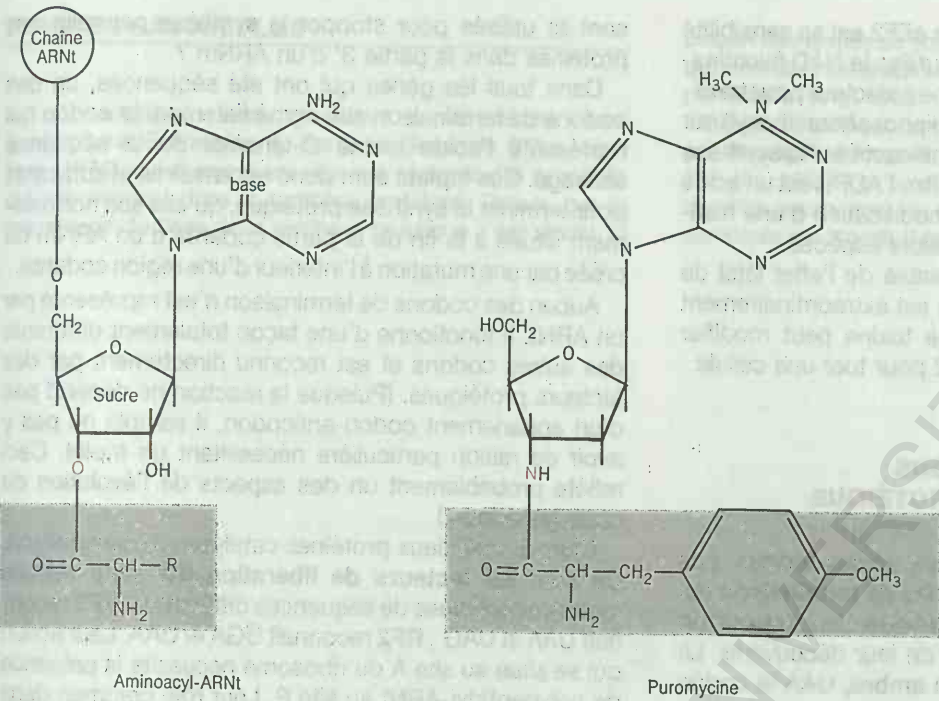


Figure 5.21 La puromycine imite l'aminocyl-ARNt car elle ressemble à un acide aminé aromatique (partie ombrée) lié à une partie sucre-base.

facteur d'élongation d'exclure l'autre, s'effectue par l'intermédiaire d'un effet sur la conformation générale du ribosome ou par une compétition directe pour la liaison à des sites chevauchants. Le fait que chaque facteur soit libéré avant que le suivant se lie assure un déroulement ordonné de la synthèse protéique.

Les deux facteurs ont besoin du GTP pour se lier au ribosome mais ne l'hydrolysent qu'ultérieurement. Puisque ni l'un ni l'autre n'utilisent le GDP pour s'unir au ribosome, la forme triphosphorylée pourrait être nécessaire à l'acquisition par le facteur de sa conformation correcte. Grâce à ce mécanisme, les facteurs n'accèdent

au ribosome qu'en compagnie du GTP qu'ils utiliseront plus tard pour accomplir leur fonction. L'hydrolyse du GTP fournit l'énergie nécessaire au changement de conformation du ribosome ; avec EF-Tu ce GTP est nécessaire à l'activation de l'aminocyl-ARNt au site A ; avec EF-G il est nécessaire au mouvement du ribosome.

L'homologue eucaryote de EF-G est la protéine eEF2 qui semble fonctionner d'une manière identique, c'est-à-dire comme une translocase dépendante de l'hydrolyse du GTP. Son action est aussi inhibée par l'acide fusidique. Un complexe stable eEF2-GTP peut être isolé ; il peut se lier aux ribosomes, ce qui entraîne l'hydrolyse du GTP.

Tableau 5.7 La translocation nécessite un facteur d'élongation.

Facteur	Codé par	Masse	Molécules /Cellule	Fonction	Inhibiteur
EF-G	fus	77 444	20 000	lie ribosome et GTP	Acide fusidique

Une caractéristique originale de eEF2 est sa sensibilité à la toxine diphthérique. La toxine utilise le NAD (nicotinamide adénine dinucléotide) comme cofacteur pour transférer une partie ADPR (adénosyl diphosphate ribosyl) sur eEF2. Le conjugué ADPR-eEF2 est inactif sur la synthèse protéique. La cible sur laquelle se fixe l'ADPR est un acide aminé inhabituel, produit par la modification d'une histidine, commun aux eEF2 de plusieurs espèces.

L'ADP ribosylation est responsable de l'effet létal de la toxine diphthérique. La réaction est extraordinairement efficace : une seule molécule de toxine peut modifier suffisamment de molécules eEF2 pour tuer une cellule.

TERMINAISON : TROIS CODONS ARRÊTENT LA SYNTHÈSE PROTÉIQUE

Seuls 61 triplets sont affectés aux acides aminés. Les trois autres triplets sont des **codons de terminaison** qui marquent la fin de la synthèse protéique. Leur appellation accidentelle résulte de l'histoire de leur découverte. Le codon UAG est appelé le codon **ambre**, UAA le codon **ocre** et le codon UGA est parfois désigné sous le nom de codon **opale**.

La nature de ces trois codons fut à l'origine mise en évidence par une technique génétique qui distinguait deux types de mutation ponctuelle.

- La mutation ponctuelle qui remplace un codon spécifique d'un acide aminé par un autre codon correspondant à un autre acide aminé est appelée mutation **faux-sens**. Un acide aminé en remplace un autre dans la protéine ; l'effet sur l'activité de la protéine dépend du site de la mutation et de la nature du nouvel acide aminé.
- Quand une mutation ponctuelle fait apparaître un codon de terminaison, la synthèse protéique s'achève **prématurément** au codon muté. Il est probable que la protéine formée sera dénuée d'activité biologique puisqu'incomplète dans la cellule mutante. Une telle mutation est une mutation **non-sens**.

(On désigne quelquefois ces triplets sous le nom de codons *non-sens*, appellation tout à fait inappropriée puisque ces codons ont un sens, quoique désagréable, pour un gène mutant.)

L'analyse des mutants non-sens a montré qu'à lui seul, un des trois codons UAG, UAA ou UGA peut arrêter la synthèse protéique dans un gène. Les mêmes signaux

sont-ils utilisés pour stopper la synthèse naturelle des protéines dans la partie 3' d'un ARNm ?

Dans tous les gènes qui ont été séquencés, un des codons de terminaison suit immédiatement le codon qui représente l'acide aminé C-terminal de la séquence sauvage. Ces triplets sont donc nécessaires et suffisants pour terminer la synthèse protéique, qu'elle soit normalement située à la fin de la partie codante d'un ARNm ou créée par une mutation à l'intérieur d'une région codante.

Aucun des codons de terminaison n'est représenté par un ARNt. Il fonctionne d'une façon totalement différente des autres codons et est reconnu directement par des facteurs protéiques. (Puisque la réaction ne dépend pas d'un appariement codon-anticodon, il semble ne pas y avoir de raison particulière nécessitant un triplet. Ceci reflète probablement un des aspects de l'évolution du code génétique.)

Chez *E. coli*, deux protéines catalysent la terminaison. Ce sont les **facteurs de libération** (RF pour release factor), spécifiques de séquences différentes. **RF1** reconnaît UAA et UAG ; **RF2** reconnaît UGA et UAA. Leur action qui se situe au site A du ribosome nécessite la présence de polypeptidyl-ARNt au site P. Leur rôle commun dans la terminaison apparaît à travers la similitude de leur séquence ; chacun a une masse d'environ 36 000 daltons, un tiers de leurs acides aminés sont identiques. Sans doute y eut-il à l'origine un seul facteur de libération qui reconnaissait tous les codons de terminaison, puis une évolution vers les deux facteurs spécifiques de codons particuliers.

Dans les systèmes eucaryotes, un seul facteur de libération existe, **eRF**. Sa liaison au ribosome dépend de la présence du GTP (ce dernier n'intervient pas chez les bactéries). Le GTP est sans doute clivé une fois la terminaison accomplie ; son hydrolyse peut être nécessaire à la dissociation eEF-ribosome.

La réaction de terminaison est constituée par la libération du polypeptide complet du dernier ARNt, de l'expulsion de l'ARNt du ribosome et de la dissociation du ribosome de l'ARNm. Le clivage entre l'ARNt et le polypeptide pourrait s'effectuer de la même façon que le transfert habituel entre un peptidyl-ARNt et un aminoacyl-ARNt au cours de l'élongation ; le facteur de libération détourne peut-être la réaction. Le mécanisme de séparation entre le ribosome et l'ARNm n'a pas encore été élucidé. Il pourrait s'agir d'un changement de conformation déclenché par le facteur RF, mais l'intervention d'autre(s) facteur(s) protéique(s) est également possible.

POUR EN SAVOIR PLUS

La découverte du code génétique a été décrite dans Lewin : *Gene Expression, 1*, *Bacterial Genomes* (Wiley, New York, 1974), qui donne aussi une vue générale des procédés qui interviennent dans la synthèse protéique. Depuis que le mécanisme de base de la synthèse protéique a été résolu, la

plupart des revues se sont centrées sur les rôles des facteurs qui contrôlent les réactions. Le modèle des facteurs a été décrit par Maitra et al. (*Ann. Rev. Biochem.* **51**, 869-900, 1982). L'initiation chez les eucaryotes et sa régulation ont été analysées par Jagus, Anderson & Safer (*Prog. Nucleic Acids. Res.* **25**, 127-185, 1981). Les procédés qui interviennent dans l'initiation chez les eucaryotes, les procaryotes et les organelles ont été comparés par Kozak (*Microbiol. Rev.* **47**, 1-45, 1983).

CHAPITRE 6

L'ARN DE TRANSFERT : L'ADAPTATEUR TRADUCTIONNEL

L'ARN de transfert occupe une position clé dans la synthèse protéique puisqu'il constitue l'élément adaptateur qui accomplit la traduction de chaque codon en acide aminé. Dans l'énoncé de Crick, l'ARNt est le moyen utilisé par la nature pour transférer l'information portée par l'acide nucléique en une protéine active. Si les ARNt sont présents dans une multitude de réactions dans lesquelles ils doivent avoir en commun certaines caractéristiques, ils doivent néanmoins se distinguer les uns des autres. La particularité essentielle qui leur confère cette capacité est le pouvoir de l'ARN d'adopter une structure tertiaire spécifique.

Si l'on revient à l'action finale de l'ARNt dans la synthèse protéique, tous les ARNt peuvent occuper les sites A et P du ribosome, associés à une extrémité à l'ARNm via l'appariement codon-anticodon et à l'autre extrémité au polypeptide à transférer. Afin que les sites A et P accueillent tous les ARNt, l'ensemble doit se conformer aux mêmes obligations générales de taille et de forme.

De la même façon, tous les ARNt (sauf l'initiateur) partagent la même possibilité d'être reconnus par les facteurs de traduction (EF-Tu ou eEF1) pour se lier au

ribosome. L'ARNt initiateur est, lui, reconnu par IF2 ou eIF2. Donc, si les ARNt doivent avoir des propriétés communes pour réagir avec les facteurs d'élongation, seul l'ARNt initiateur doit en être dépourvu ou en posséder une opposée.

D'autre part, il doit y avoir des différences critiques entre de petits groupes d'ARNt. Habituellement, chaque acide aminé est représenté par plus d'un ARNt. Plusieurs ARNt représentant le même acide aminé sont appelés **ARNt isoaccepteurs**. *Un groupe d'ARNt isoaccepteurs est obligatoirement chargé par la seule aminoacyl-ARNt synthétase spécifique de leur acide aminé.* Les ARNt isoaccepteurs sont donc dotés de caractères communs qui permettent à l'enzyme de les repérer au sein des autres ARNt. L'ensemble des ARNt se répartit en 20 groupes d'isoaccepteurs ; chaque groupe est capable de s'identifier à sa synthétase propre.

Nous savons encore peu de choses des éléments qui permettent aux ARNt d'identifier les protéines avec lesquelles ils réagissent. Les études préliminaires visant à faire apparaître des caractères communs se concentrent sur la recherche de courtes séquences de nucléotides partagées par tous les ARNt ou par un groupe

particulier d'ARNt. Cependant, bien que certains éléments de la séquence primaire de l'ARNt soient hautement conservés, sans doute essentiels pour une ou plusieurs de ses fonctions, il n'a pas été possible de corrélérer ces séquences avec le mode de reconnaissance de la protéine.

Les caractères communs (ou distinctifs) des ARNt peuvent aller au-delà des structures primaire et secondaire et être portés au moins en partie par la structure tertiaire. L'analyse détaillée du rôle de la structure tertiaire nécessite la comparaison de plusieurs si ce n'est de tous les ARNt. Il s'agit d'une lourde tâche, puisqu'aujourd'hui seules quelques molécules d'ARNt ont été analysées en ce qui concerne leur structure tertiaire.

LA FEUILLE DE TRÈFLE UNIVERSELLE

Les séquences de plusieurs centaines d'ARNt isolés d'une grande variété de bactéries et d'eucaryotes ont été déterminées. Toutes se conforment à la même structure secondaire générale. Chaque séquence d'ARNt peut s'écrire sous la forme d'une **feuille de trèfle** maintenue par l'appariement de courtes régions complémentaires (voir Figure 5.2).

La forme générale de la feuille de trèfle est illustrée dans la **Figure 6.1**. Les quatre **bras** principaux sont désignés en fonction de leur structure ou de leur rôle. Le **bras accepteur** forme une échelle dont l'extrémité est une séquence non appariée dont un groupe OH libre 2' ou 3' est aminoacylé. Les autres bras sont également constitués de bases appariées formant une **échelle** et des **boucles** non appariées. Le **bras T ψ C** est ainsi désigné en raison de la présence de cette séquence de nucléotides, le **bras de l'anticodon** contient toujours le triplet de l'anticodon au centre de la boucle et le **bras D** doit son nom au fait qu'il contient la dihydro-uridine (ψ et D représentent respectivement la pseudo-uridine et la dihydro-uridine, deux bases « inhabituelles » de l'ARNt qui seront étudiées plus tard).

Le système de numération de l'ARNt illustre bien la constance de la structure. Les positions sont numérotées en allant du 5' vers le 3', selon la structure la plus commune de l'ARNt constituée de 76 nucléotides. La longueur de l'ARNt varie de 74 à 95 bases. Ces différences de longueur résultent de modifications de structure qui n'interviennent que sur deux bras.

Dans la boucle D, les variations peuvent concerner jusqu'à 4 résidus. Les nucléotides supplémentaires relatifs à la structure la plus commune sont notés 17:1 (entre

17 et 18) et 20:1 et 20:2 (entre 20 et 21). Cependant la boucle D la plus courte peut être dépourvue du résidu 17 ainsi que des trois autres résidus.

La plus grande variété au niveau de l'ARNt réside dans ce que l'on appelle le **bras supplémentaire** situé entre le bras T ψ C et celui de l'anticodon. Selon la nature du bras supplémentaire, les ARNt se répartissent en deux catégories. Les **ARNt de classe 1** ont un bras petit supplémentaire constitué de 3 à 5 bases. Ils représentent 75 % des ARNt. Les **ARNt de classe 2** ont un bras supplémentaire plus long — il peut même être le plus long de l'ARNt — avec 13 à 21 bases et environ 5 bases appariées. Les bases supplémentaires sont numérotées de 47:1 à 47:16.

L'appariement de bases complémentaires qui maintient la structure secondaire est pratiquement invariable. L'image de la feuille de trèfle parcourue dans le sens des aiguilles d'une montre fait apparaître qu'il y a toujours 7 paires de bases dans le bras accepteur, 5 dans le bras T ψ C, 5 dans le bras anticodon et habituellement 3 (parfois 4) dans le bras D. Pour un ARNt donné, les bases s'apparient selon le mode conventionnel A-U et G-C mais des appariements de type G-U, G- ψ ou A- ψ peuvent se former. D'une moins grande stabilité, ces liaisons permettent tout de même à l'ARN d'adopter une structure en double hélice.

Si l'on compare les séquences des ARNt, les bases qui occupent certaines positions sont **invariantes** (ou **conservées**) ; presque toujours, une base particulière est trouvée dans cette position. En fait, avec le nombre d'ARNt séquencés croissant, les positions qui semblaient totalement invariables présentent quelques exceptions. Ainsi la description d'une position conservée signifie que la base spécifique est présente dans plus de 90 à 95 % des ARNt. Les exceptions sont parfois isolées, elles peuvent aussi faire partie de groupes qui représentent une particularité d'une cellule.

Des positions décrites en tant que **semi-invariantes** (ou **semi-conservées**) semblent restreintes à un type de bases (purine contre pyrimidine) mais l'une ou l'autre base de ce type peut être présente.

LA STRUCTURE TERTIAIRE EST EN FORME DE L

La représentation en feuille de trèfle de la structure secondaire de l'ARNt ne permet pas de rendre compte de la structure tertiaire, en fait plutôt compacte. L'étude de la structure tertiaire par cristallographie aux rayons X

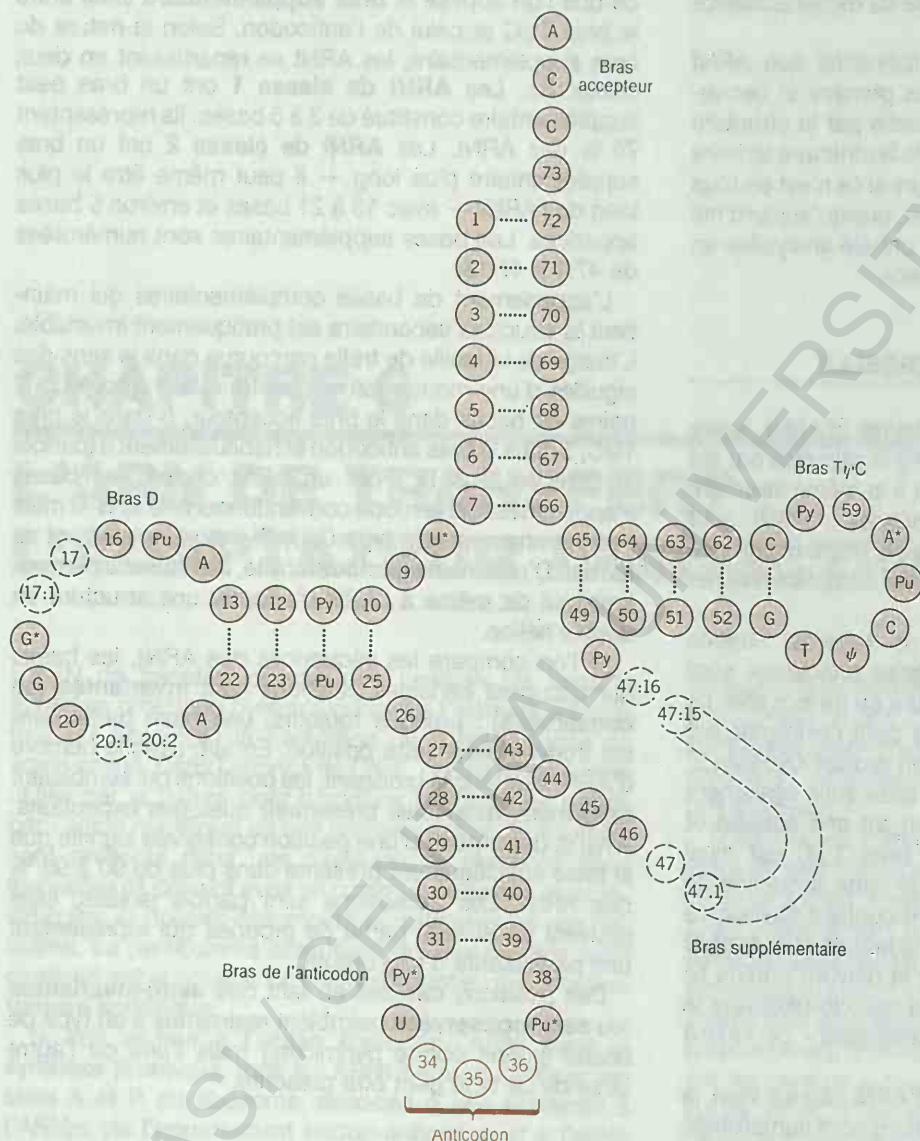


Figure 6.1.

L'ARN de transfert forme une feuille de trèfle.

Les cercles indiquent les positions des bases. Les nombres qui apparaissent dans les cercles ombrés indiquent les positions qui sont toujours présentes mais qui peuvent être occupées par n'importe quelle base. Quand une position est toujours occupée par une base particulière, la base remplace le nombre. Pour les bases invariantes, la base courante est indiquée ; pour les bases semi-invariantes, Py et Pu indique la présence soit d'une pyrimidine, soit d'une purine. Un astérisque indique que la base est modifiée, mais que la forme de la modification peut varier. Les cercles ouverts soulignent les positions présentes uniquement dans quelques ARNt. Les pointillés indiquent l'appariement des bases.

nécessite l'obtention de cristaux d'ARNt. Cette entreprise difficile n'a été réalisée que pour quelques ARNt, isolés pour la plupart de levure. Toutefois les similarités de structure suggèrent qu'un motif commun de structure tertiaire est respecté par tous les ARNt, bien que chacun possède ses variations propres.

La double hélice de la structure secondaire est maintenue dans la structure tertiaire, mais son organisation dans l'espace crée principalement deux doubles hélices formant un angle droit, comme le montre la **Figure 6.2**. La tige du bras accepteur et celle du bras T ψ C constituent une double hélice continue interrompue une seule fois, de même pour le bras D et le bras de l'anticodon. La région entre les doubles hélices où se forme l'angle du « L » contient les boucles T ψ C et D. Ainsi, l'extrémité d'un bras du L porte l'acide aminé et la boucle anticodon forme l'autre extrémité.

Un modèle moléculaire de l'ARNt^{Phe} de levure est illustré par la **Figure 6.3**. Ce modèle est considéré comme la forme générale de la structure tertiaire des ARNt. Chaque ARNt présente des différences ponctuelles affectant une région ou une autre ; ceci permet de concilier le fait que tous les ARNt doivent avoir une forme identique bien qu'il soit possible de reconnaître des différences parmi eux. Par exemple, les deux axes de l'ARNt^{Asp}

forment un angle légèrement plus grand donnant à la molécule une conformation plus ouverte.

La structure tertiaire est due à des liaisons hydrogène qui font intervenir pour la plupart des bases non appariées de la structure secondaire. Les liaisons de la feuille de trèfle sont appelées **liaisons H secondaires**, les liaisons supplémentaires de la structure tertiaire sont les **liaisons H tertiaires**. Parmi ces dernières, beaucoup font intervenir des bases invariantes et semi-invariantes, ce qui explique leur conservation et suggère aussi que la forme générale de la structure tertiaire est commune à tous les ARNt. Aucune de ces interactions n'est universelle, mais elles sont la marque du schéma *général* d'élaboration de la structure de l'ARNt.

La diffraction aux rayons X permet de visualiser une organisation stable de la structure de la molécule. En solution ou associée à des protéines, une certaine flexibilité peut être mise en évidence. Dans ces deux cas la conformation de l'ARNt peut varier, mais aucun lien entre ces changements et la structure cristalline n'a pu être établi.

La structure amène une conclusion générale à propos du rôle de l'ARNt. *Ce sont les deux sites les plus éloignés l'un de l'autre qui exercent des rôles spécifiques*. L'acide aminé est éloigné le plus possible de l'anticodon, ce qui

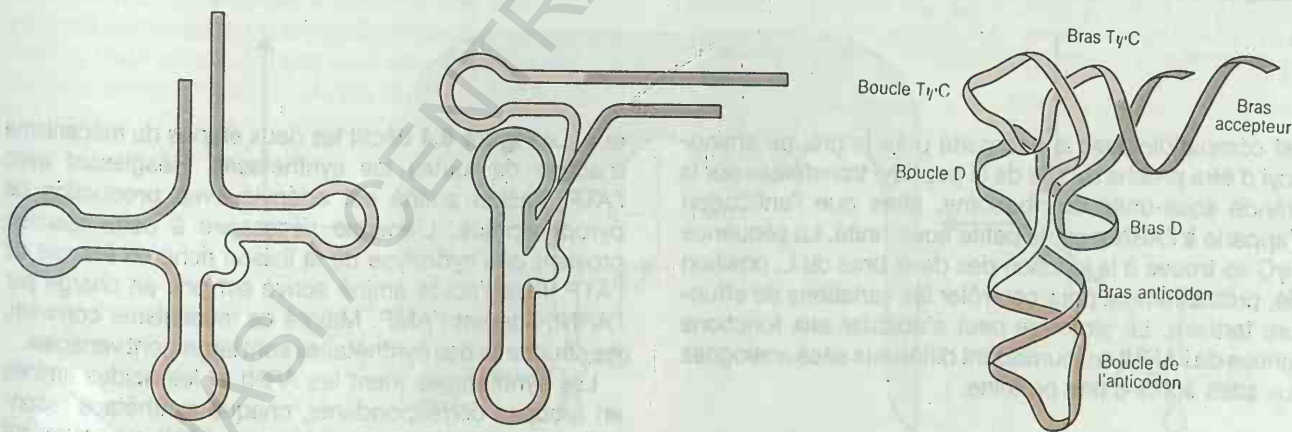


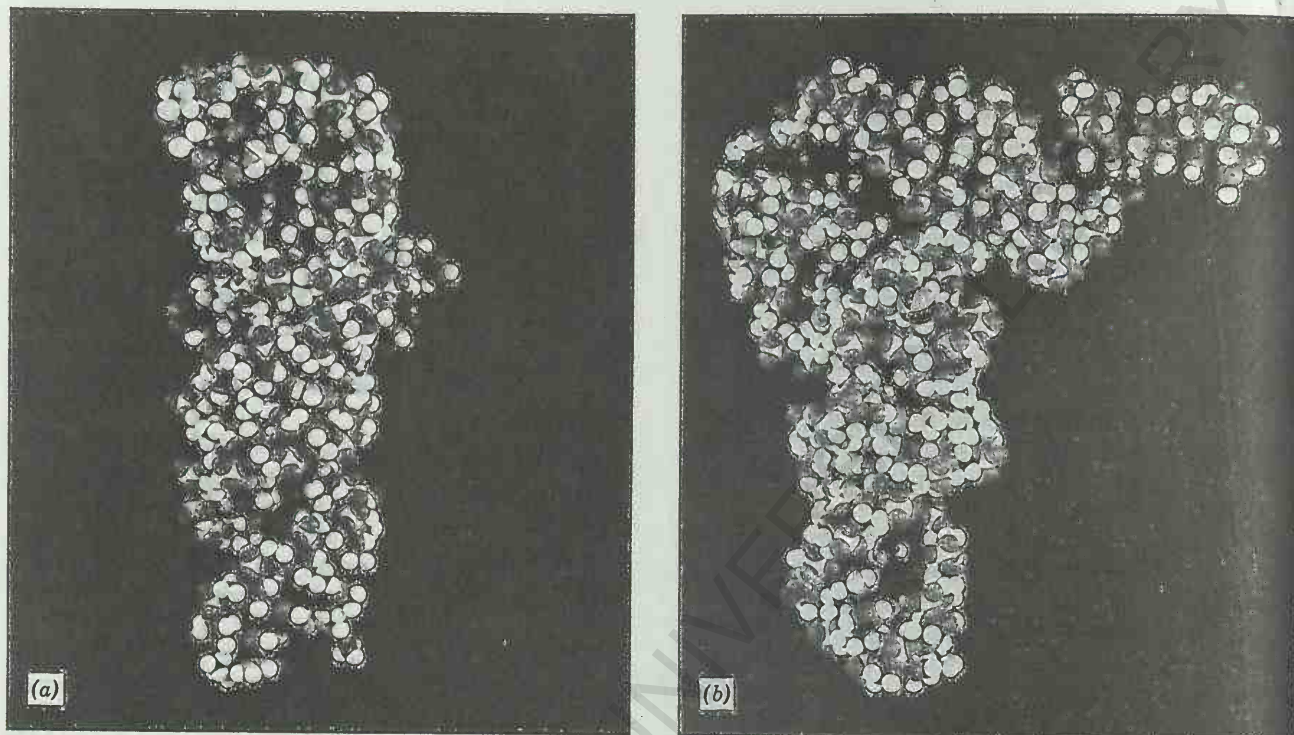
Figure 6.2
L'ARN de transfert se replie en une structure tertiaire compacte en forme de L.

A gauche. Une feuille de trèfle avec les bras individualisés.

Au centre. Une projection en deux dimensions montrant comment les régions en double hélice forment des angles droits.

A droite. Une représentation schématique dans laquelle l'épine dorsale peut être tracée à travers les régions de bases appariées et non appariées.

Ceci est la structure de l'ARNt^{Phe} de levure, mais les autres sont similaires.

**Figure 6.3**

Un modèle de représentation dans l'espace montre que la structure tertiaire de l'ARNt est compacte.

Une rotation de 90° a été effectuée entre les deux images de l'ARNt^{Phe}. La vue de droite correspond à la partie droite de la figure précédente.

Photographie aimablement prêtée par S. H. Kim.

est compatible avec la nécessité pour le groupe aminoacyl d'être proche du site de la peptidyl transférase sur la grande sous-unité du ribosome, alors que l'anticodon s'apparie à l'ARNm sur la petite sous-unité. La séquence T ψ C se trouve à la jonction des deux bras du L, position clé, probablement pour contrôler les variations de structure tertiaire. La structure peut s'adapter aux fonctions variées de l'ARNt en fournissant différents sites analogues aux sites actifs d'une protéine.

COMMENT LES SYNTHÉTASES RECONNAISSENT-ELLES LES ARNt ?

Les acides aminés entrent dans la synthèse protéique par l'intermédiaire des aminoacyl-ARNt synthétases qui sont les interfaces pour la connexion avec l'acide nucléi-

que. La **Figure 6.4** décrit les deux étapes du mécanisme d'action de toutes les synthétases. Réagissant avec l'ATP, l'acide aminé est adénylé avec production de pyrophosphate. L'énergie nécessaire à cette réaction provient de l'hydrolyse de la liaison riche en énergie de l'ATP. Puis l'acide aminé activé est pris en charge par l'ARNt, libérant l'AMP. Malgré ce mécanisme commun, les structures des synthétases sont largement variables.

Les synthétases trient les ARNt et les acides aminés en groupes correspondants, chaque synthétase reconnaît un seul acide aminé et tous les ARNt qui pourraient le prendre en charge. (Il peut y avoir plusieurs ARNt dans ce cas. En plus des ARNt nécessaires pour répondre aux codons synonymes, des espèces multiples d'ARNt peuvent parfois réagir avec le même codon). Les ARNt convenablement reconnus par une synthétase sont décrits comme ses ARNt **apparentés**.

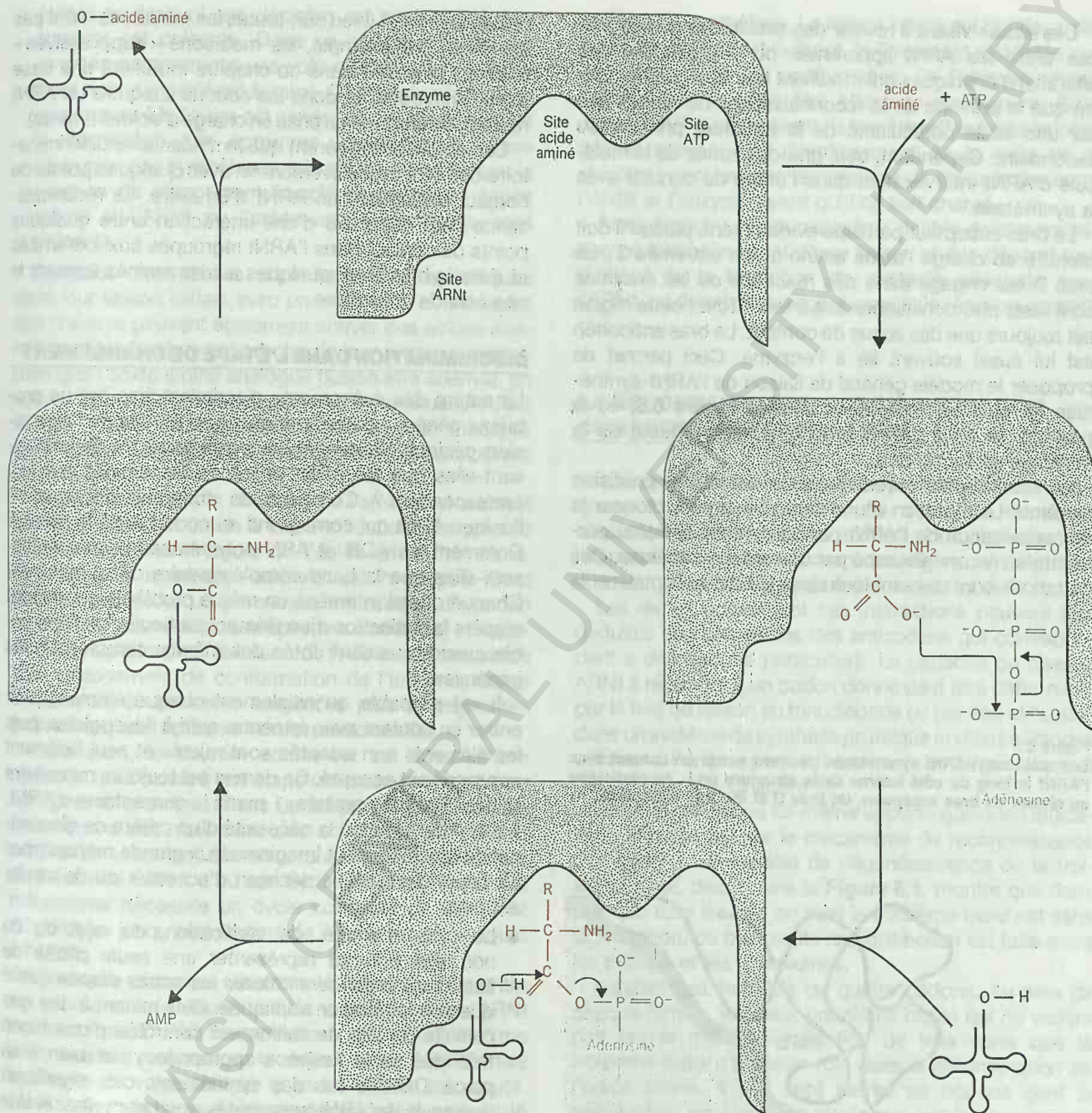


Figure 6.4
Une aminoacyl-ARNt synthétase charge l'ARNt avec un acide aminé.

L'enzyme présente différents sites pour l'ARNt, un acide aminé et l'ATP. La première réaction requiert l'acide aminé et l'ATP ; elle produit un aminoacyl-AMP. Puis l'ARNt est lié et l'acide aminé lui est transféré. Quand l'ainoacyl-ARNt est libéré, l'enzyme est prête pour un autre cycle.

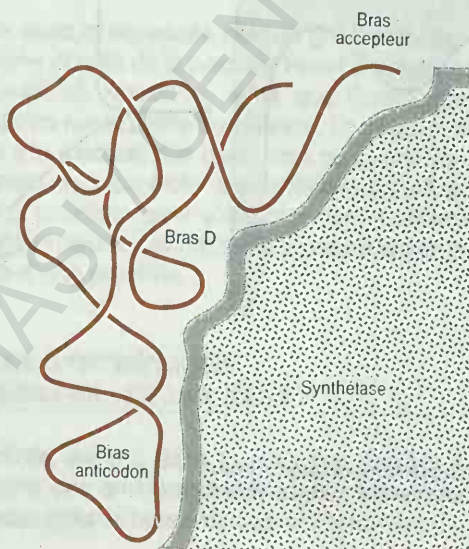
Des essais visant à révéler des similitudes de séquences entre les ARNt apparentés ou à provoquer des altérations chimiques qui modifient leur charge ont montré que le principe de la reconnaissance ne repose pas sur une seule particularité de la structure primaire ou secondaire. Cependant, trois grandes zones de la molécule d'ARNt interviennent dans l'étape de contact avec la synthétase.

Le bras accepteur participe évidemment, puisqu'il doit prendre en charge l'acide aminé à son extrémité 3'. Le bras D est engagé dans des réactions où les enzymes sont liées photochimiquement à leur ARNt ; cette région est toujours une des zones de contact. Le bras anticodon est lui aussi souvent lié à l'enzyme. Ceci permet de proposer le modèle général de liaison de l'ARNt-synthétase tel qu'il est représenté dans la **Figure 6.5**, où la protéine se lie à l'ARNt le long du côté interne de la molécule en forme de L.

La réaction de reconnaissance est d'une précision extrême. La mutation d'une seule base peut bloquer la reconnaissance de l'ARNt par sa synthétase et/ou permettre sa reconnaissance par une autre synthétase. Ces mutations sont rassemblées dans le bras accepteur et le

Figure 6.5

Les aminoacyl-ARNt synthétases peuvent entrer en contact avec l'ARNt le long du côté interne de la structure en L, en particulier au niveau du bras accepteur, du bras D et du bras anticodon.



bras anticodon * (bien sûr, toutes les mutations n'ont pas cet effet ; par exemple, les mutations « suppressives » étudiées plus tard dans ce chapitre modifient une base dans l'anticodon, et donc les codons auxquels un ARNt répond, sans altérer sa prise en charge d'acides aminés).

Ces résultats suggèrent que le mécanisme discriminatoire doit être excessivement fin avec quelques points de contact distinguant un ARNt d'un autre. La reconnaissance peut dépendre d'une interaction entre quelques points de contact dans l'ARNt regroupés aux extrémités et dans le bras D, et quelques acides aminés formant le site actif de la protéine.

DISCRIMINATION DANS L'ÉTAPE DE CHARGEMENT

La nature des événements discriminatoires est un problème général soulevé par plusieurs étapes de l'expression génique. Comment les synthétases ne reconnaissent-elles que les ARNt et les acides aminés qui leur correspondent ? Comment un ribosome distingue-t-il l'unique ARNt qui correspond au codon dans le site A ? Comment les ADN et ARN polymérases ne reconnaissent-elles que la base complémentaire de la matrice ? Chaque question amène un même problème : comment s'opère la distinction d'un élément particulier de l'ensemble quand tous sont dotés des mêmes caractéristiques générales.

Il est probable qu'initialement chaque élément peut entrer en contact avec le centre actif à l'aveuglette, puis les éléments non adaptés sont rejetés et seul l'élément approprié est accepté. Ce dernier est toujours minoritaire (1 des 20 acides aminés, 1 parmi la quarantaine d'ARNt, 1 des 4 bases) d'où la nécessité d'un critère de discrimination strict. On peut imaginer deux grands mécanismes au cours desquels la décision d'accepter ou de rejeter est prise.

- Le cycle d'entrée, de vérification, de rejet ou de non rejet pourrait représenter une seule phase de liaison qui *précéderait toutes les autres étapes* quelle que soit la réaction impliquée. Cela revient à dire que l'affinité du site de liaison est contrôlée d'une façon telle que seule l'espèce appropriée y est bien à sa place. Dans le cas des synthétases, cela signifierait que seuls les ARNt apparentés pourraient réaliser une fixation stable au site.
- Dans l'autre alternative, la réaction peut procéder par

* N.d.T. : Un premier déchiffrement du « code » de correspondance entre l'ARNt et l'aminocyl-ARNt synthétase a été rapporté en mai 1988. Une paire de bases G-U située à la troisième position de la partie double-brin du bras accepteur de l'ARN^{Ala} suffit à conférer à l'ARN^{Phe} auquel elle a été transférée la propriété d'être chargé par de l'alanine, et non par de la phénylalanine.

étapes après quoi une décision est prise si l'espèce correcte est présente. Dans ce cas, la réaction est inversée ou emprunte une voie de dérivation et l'élément non approprié est éliminé. Ce type de vérification postérieure à la liaison est généralement décrit comme la **correction sur épreuves** (proofreading). Dans l'exemple des synthétases, elle sous-entendrait que la réaction de chargement franchirait certaines étapes même si l'ARNt ou l'acide aminé inadaptés, sont présents.

Quelques synthétases sont hautement spécifiques dans leur liaison initiale avec un seul acide aminé, alors que d'autres peuvent également activer des acides aminés dont la structure est proche de leur propre substrat. Bien que l'acide aminé analogue puisse être adénylé, en aucun cas un tel dérivé ne peut former un aminoacyl-ARNt stable. A quel moment l' aminoacyl-adénylé inadapté est-il rejeté ? La **Figure 6.6** résume les choix qui s'offrent à l'enzyme.

Il est possible que la structure de l' aminoacyl adénylé soit simplement soumise à une vérification plus précise que ne l'a été la structure de l'acide aminé lui-même. Puis le faux aminoacyl adénylé est hydrolysé. Ceci pourrait constituer un processus intrinsèque d'examen plus précis (**correction sur épreuve cinétique**) ; il interviendrait également si la liaison de l'ARNt était inductrice d'un changement de conformation de l'enzyme, lequel entraînerait de nouvelles exigences plus astreignantes (**correction sur épreuve conformationnelle**).

Cependant, le courant actuel est en faveur d'une sélection qui s'opère à une étape ultérieure par **correction sur épreuve chimique**. Dans un tel procédé, le mauvais acide aminé est réellement pris en charge par l'ARNt, puis reconnu en tant que structure inappropriée au site de liaison de l'ARNt d'où son hydrolyse et son rejet. Ce mécanisme nécessite un cycle continu de liaison et d'hydrolyse jusqu'au transfert de l'acide aminé adéquat à l'ARNt.

L'exemple classique illustrant la discrimination entre les acides aminés dépendant de la présence de l'ARNt est fourni par l'ARNt Ile-synthétase d'*E. coli*. L'enzyme peut prendre en charge la valine avec l'AMP, mais hydrolyse le valyl-adénylé quand l'ARNt^{Ile} est ajouté. L'estimation de l'erreur globale dépend de la spécificité de chaque étape comme cela est résumé dans le **Tableau 6.1**. Le taux d'erreur global de $1,5 \times 10^{-5}$ est inférieur au taux mesuré de substitution de l'isoleucine par la valine (dans la globine de lapin), qui est de $2,5 \times 10^{-4}$. Ainsi un chargement défectueux est-il responsable d'une petite fraction d'erreurs qui apparaissent réellement.

La **Figure 6.7** illustre les réactions de liaison entre

l'ARNt et la synthétase. La liaison initiale est rapide ; puis, si l'ARNt correct est présent, la liaison est stabilisée par un changement de conformation de l'enzyme. Ainsi, l' aminoacylation peut être effectuée rapidement. Si le mauvais ARNt est présent, la conformation reste inchangée et provoque le ralentissement de la suite de la réaction ; ceci favorise les chances de dissociation entre l'ARNt et l'enzyme avant qu'il ne soit chargé.

Ainsi, il peut y avoir un système double de discrimination où l'établissement d'une liaison stable dépend tout d'abord de la formation de contacts adéquats avec l'ARNt, ce dernier influençant ensuite la réaction catalytique.

LA RECONNAISSANCE CODON-ANTICODON COMPORTE UNE OSCILLATION

Le rôle de l'ARNt dans la synthèse protéique est accompli quand il reconnaît le codon au site A du ribosome. L'interaction entre le codon et l'anticodon s'effectue par l'appariement de bases suivant des règles qui s'étendent au-delà des habituels appariements A-U et G-C.

Les règles gouvernant ces interactions peuvent être déduites des séquences des anticodons qui correspondent à des codons particuliers. La capacité de chaque ARNt à répondre à un codon donné peut être déterminée par le test de liaison au trinuécléotide ou par son utilisation dans un système de synthèse protéique *in vitro* (méthodes utilisées à l'origine pour déterminer le code génétique et décrites dans le Chapitre 5).

Le code génétique lui-même apporte quelques indications importantes sur le mécanisme de reconnaissance des codons. Le modèle de dégénérescence de la troisième base, décrit dans la **Figure 6.8**, montre que dans presque tous les cas ou bien la troisième base est sans signification, ou bien seule une distinction est faite entre les purines et les pyrimidines.

Il existe huit **familles** de quatre codons. Au sein de chaque famille, les deux premières bases qui ne varient pas ont la même signification de telle sorte que la troisième base n'a aucun rôle dans la détermination de l'acide aminé. Il y a sept **paires** de codons dont la signification est identique quelle que soit la base pyrimidique présente ; enfin, dans cinq paires de codons, le même acide aminé est codé quelle que soit la base purique.

Il est encore plus significatif de regarder le code en sens inverse. *Dans trois cas seulement, la présence d'une base particulière en troisième position confère une seule signification* : AUG (pour la méthionine), UGG (pour le

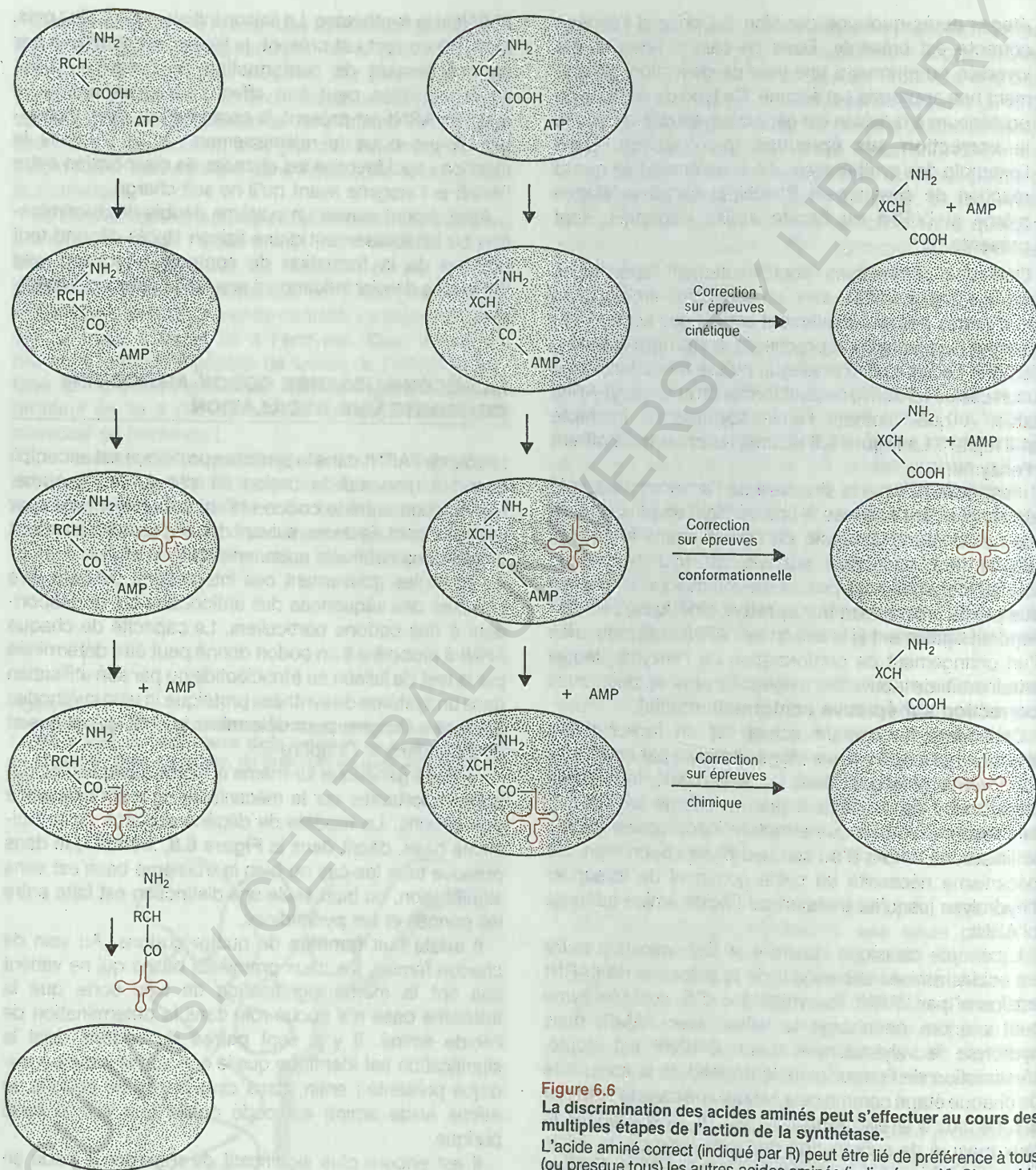


Figure 6.6

La discrimination des acides aminés peut s'effectuer au cours des multiples étapes de l'action de la synthétase.

L'acide aminé correct (indiqué par R) peut être lié de préférence à tous (ou presque tous) les autres acides aminés (indiqués par X). Si un acide aminé incorrect est lié, il peut être soumis à une correction sur épreuves cinétique, grâce à un changement de conformation induit par l'ARNt, ou chimique via un transfert abortif d'ARNt.

Tableau 6.1

La précision de chargement de l'ARNt^{le} par sa synthétase est sous la dépendance d'un contrôle d'erreur en deux étapes.

Étape	Fréquence d'erreur
Activation de la valine en Val-AMP	1/225
Libération du Val-ARNt ^{le}	1/270
Taux global d'erreur	$1/225 \times 1/270 = 1/60\ 000$

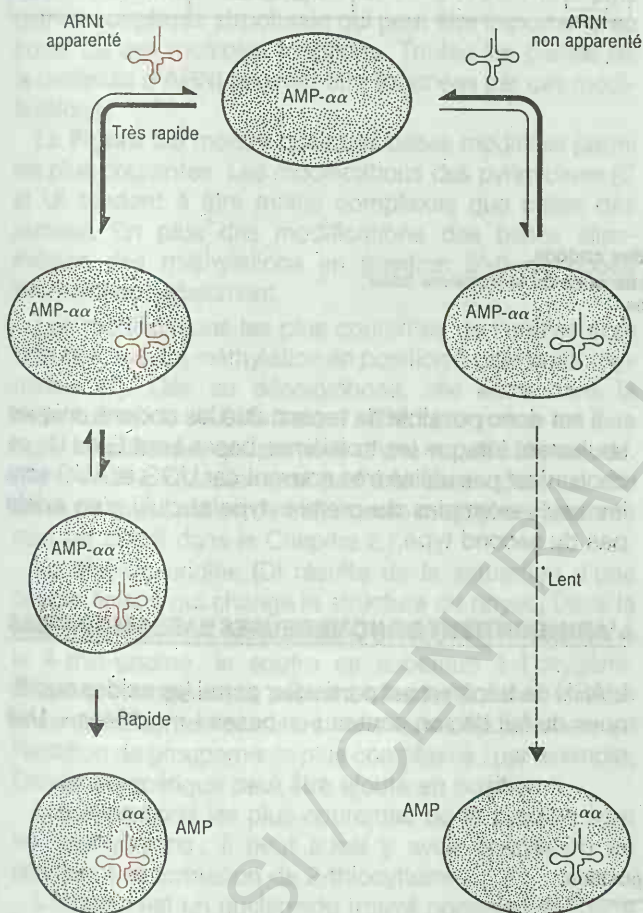


Figure 6.7

La reconnaissance de l'ARNt correct par la synthétase subit un contrôle en deux étapes.

Premièrement, l'enzyme a une affinité plus grande pour son ARNt apparenté, dont la liaison est stabilisée par une transition conformationnelle dans l'enzyme.

Deuxièmement, l'aminoacylation de l'ARNt incorrect est très lente, ce qui ultérieurement augmente la probabilité d'une dissociation de l'ARNt.

tryptophane), et UGA (non-sens). Ainsi, C et U en troisième position n'ont jamais une signification unique et A ne signifie jamais un seul acide aminé.

Comme l'anticodon est complémentaire du codon, c'est la *première* base de la séquence de l'anticodon, écrite selon la convention du 5' vers le 3', qui s'apparie avec la *troisième* base de la séquence du codon, écrite selon la même convention. Ainsi, la combinaison

Codon	5'	ACG	3'
Anticodon	3'	UGC	5'

est habituellement écrite sous la forme codon ACG/anticodon CGU, la séquence de l'anticodon devant être lue en sens inverse en raison de sa complémentarité avec le codon.

Pour éviter toute confusion, nous retiendrons la convention habituelle dans laquelle toutes les séquences sont écrites 5'-3', mais nous indiquerons les séquences anticodon par une flèche dirigée en arrière afin de rappeler la relation avec le codon. Ainsi, la paire codon/anticodon représentée ci-dessus s'écrira ACG et $\overleftarrow{\text{CGU}}$ respectivement.

Chaque codon exige-t-il son propre ARNt avec un anticodon complémentaire ? Ou bien, quelques ARNt peuvent-ils répondre aux deux membres d'une paire de codons et à tous (ou quelques-uns) les membres d'une famille de codons ?

Un ARNt peut souvent reconnaître plus d'un codon. Ainsi, la base en première position de l'anticodon doit pouvoir s'associer aux bases alternatives occupant la troisième position du codon. L'appariement de bases dans cette position ne peut se limiter aux habituelles associations G-C et A-U.

Les règles gouvernant le mode de reconnaissance constituent l'**hypothèse de l'oscillation** (wobble hypothesis), qui établit que la complémentarité des deux premières bases entre codon et anticodon suit les lois habituelles, mais qu'une certaine oscillation exceptionnelle peut apparaître en troisième position. Ce phéno-

UUU	UCU	UAU	UGU
UUC	UCC	UAC	UGC
UUA	UCA	UAA	UGA
UUG	UCG	UAG	UGG
CUU	CCU	CAU	CGU
CUC	CCC	CAC	CGC
CUA	CCA	CAA	CGA
CUG	CCG	CAG	CGG
AUU	ACU	AAU	AGU
AUC	ACC	AAC	AGC
AUA	ACA	AAA	AGA
AUG	ACG	AAG	AGG
GUC	GCU	GAU	GGU
GUA	GCC	GAC	GGC
GUG	GCA	GAA	GGA
	GCG	GAG	GGG

Relation avec la 3 ^e base	3 ^e bases avec la même signification	Nombre de codons
Troisième base sans signification	U, C, A, G	32 (8 familles)
Distinction entre les purines et les pyrimidines	U ou C A ou G	14 (7 paires) 12 (6 paires)
Trois sur quatre	U, C, A	3 (AUX = Ile)
Définitions uniques	G seulement	2 (AUG = Met UGG = Trp)
	A seulement	1 (UGA = term)

Figure 6.8

Les bases en troisième position influent le moins sur la signification des codons.

Les rectangles indiquent les groupes de codons pour lesquels la dégénérescence de la troisième base assure une signification identique. Ils peuvent être divisés en quatre classes.

mène provient de la conformation de la boucle anticodon de l'ARNt qui autorise une flexibilité inhabituelle de la première base de l'anticodon.

Le mode de reconnaissance de la troisième base du codon permet l'appariement entre G et U en plus des paires habituelles. Ce simple changement crée un mode d'appariement de bases dans lequel A ne peut plus avoir un seul sens dans le codon (car U qui la reconnaît doit aussi reconnaître G). Il en va de même pour C (car G qui la reconnaît doit aussi reconnaître U). Le **Tableau 6.2** résume le mode de reconnaissance.

Il est donc possible de reconnaître les codons uniques seulement lorsque les troisièmes bases sont G ou U ; ce choix n'est pas utilisé très souvent car UGG et AUG sont les seuls exemples du premier type et qu'il n'en existe pas du second type.

L'ARNt CONTIENT DE NOMBREUSES BASES MODIFIÉES

L'ARN de transfert est particulier parmi les acides nucléiques du fait de son contenu en bases « modifiées ». Une

Tableau 6.2

L'appariement codon-anticodon comporte une oscillation en troisième position.

Base en première position de l'anticodon	Base(s) reconnue(s) en troisième position du codon
U	A ou G
C	G seulement
A	U seulement
G	C ou U

base modifiée est n'importe quel cycle purique ou pyrimidique exceptions faites de A, G, C et U qui contribuent à la synthèse de tous les ARN. Toutes les autres bases sont issues de la **modification** d'une des quatre bases une fois incorporée dans la chaîne polyribonucléotidique.

Toutes les espèces d'ARN montrent un certain degré de modification, mais à l'exception de l'ARNt, dans tous les cas il ne s'agit que d'événements simples tels que l'addition de groupes méthyl. Dans l'ARNt, il existe une grande variété de modifications allant de simples méthylations jusqu'à une restructuration globale du noyau purine. Ces modifications confèrent à l'ARNt une plus grande souplesse structurale qui peut être importante au cours de ses multiples fonctions. Toutes les parties de la molécule d'ARNt peuvent être touchées par ces modifications.

La **Figure 6.9** montre quelques bases modifiées parmi les plus courantes. Les modifications des pyrimidines (C et U) tendent à être moins complexes que celles des purines. En plus des modifications des bases elles-mêmes des méthylations en position 2'-O du ribose interviennent également.

Les modifications les plus courantes de l'uridine sont bien définies. La méthylation en position 5 crée la ribothymidine (T). Liée au désoxyribose, elle entre dans la constitution de l'ADN. Dans l'ARN, la thymine est une base inhabituelle, qui provient d'une modification de U. (Dans l'ADN, l'uracile n'est pas une base courante, elle apparaît sous l'action de traitements mutagènes, comme cela est décrit dans le Chapitre 2.)

La dihydro-uridine (D) résulte de la saturation d'une double liaison qui change la structure du noyau. Dans la pseudo-uridine (ψ), les atomes C et N sont inversés. Dans la 4-thio-uridine, le soufre se substitue à l'oxygène. D'autres changements peuvent affecter l'uridine comme la combinaison de plusieurs variations déjà décrites, ou l'addition de groupements plus complexes ; par exemple, l'acide oxyacétique peut être ajouté en position 5.

Les altérations les plus courantes de la cytidine sont les méthylations ; il peut aussi y avoir acétylation en position 4 et formation de 2-thiocytidine.

L'inosine est un nucléoside trouvé normalement dans la cellule comme intermédiaire de la biosynthèse des purines. Toutefois, il n'est pas incorporé directement dans l'ARN où sa présence dépend de la modification de A. D'autres modifications de A sont constituées d'additions de groupements complexes.

Deux séries complexes de nucléotides résultent de modifications sur G. Les bases Q, telle la queuosine,

possèdent un noyau pentényl supplémentaire lié au groupe méthyl de la 7-méthylguanosine par une liaison NH. Ce noyau pentényl peut lui-même porter d'autres groupes. Les bases Y, telle la wyosine, ont un noyau supplémentaire fusionné au noyau purine lui-même ; ce noyau additif porte une longue chaîne carbonée qui peut, selon les cas, porter à nouveau d'autres groupes.

La réaction de modification implique habituellement l'altération des bases de l'ARNt ou l'adjonction d'autres groupes. La synthèse des bases Q fait exception à la règle puisqu'une enzyme particulière échange la queuosine avec un résidu guanosine de l'ARN. Cette réaction nécessite la rupture et la reconstitution de liaisons de part et d'autre du nucléoside, réaction exceptionnelle dans la synthèse des acides nucléiques.

On dénombre actuellement une cinquantaine de nucléosides modifiés dans l'ARNt. Ils sont synthétisés par des enzymes de modification ARNt-spécifiques. Le nucléoside d'origine peut être identifié à chaque position, soit en comparant la séquence de l'ARNt avec celle de son gène, soit (moins efficacement) en isolant les molécules précurseurs de plusieurs, voire de toutes les modifications. Les séquences des précurseurs montrent que les différentes modifications sont introduites au cours des différentes étapes de la maturation de l'ARNt.

Certaines modifications sont constantes dans toutes les molécules d'ARNt — par exemple, les résidus D qui donnent leur nom au bras D et la ψ trouvée dans la séquence T ψ C. En 3', l'anticodon a toujours une purine modifiée ; la modification introduite varie beaucoup.

D'autres modifications sont spécifiques d'ARNt particuliers ou de groupes d'ARNt. Par exemple, les bases wyosine sont caractéristiques de l'ARNt^{Phe} de bactéries, de levure et de mammifères. Il existe aussi d'autres modifications spécifiques d'espèces.

On ne connaît pas encore les caractéristiques reconnues par les enzymes de modification des ARNt. Quand une modification particulière apparaît dans plus d'une position de l'ARNt, ce n'est pas forcément la même enzyme qui effectue tous ces changements ; ainsi des enzymes de modifications différentes peuvent être nécessaires à la synthèse de la pseudo-uridine dans chaque position. Si le mécanisme de contrôle de spécificité des enzymes de modification n'est pas encore connu, il est clair qu'il existe plusieurs enzymes avec des spécificités variables. Certaines enzymes ne catalysent qu'une seule réaction avec un ARNt donné, d'autres peuvent utiliser plusieurs substrats. Quelques modifications nécessitent les actions successives de plusieurs enzymes.

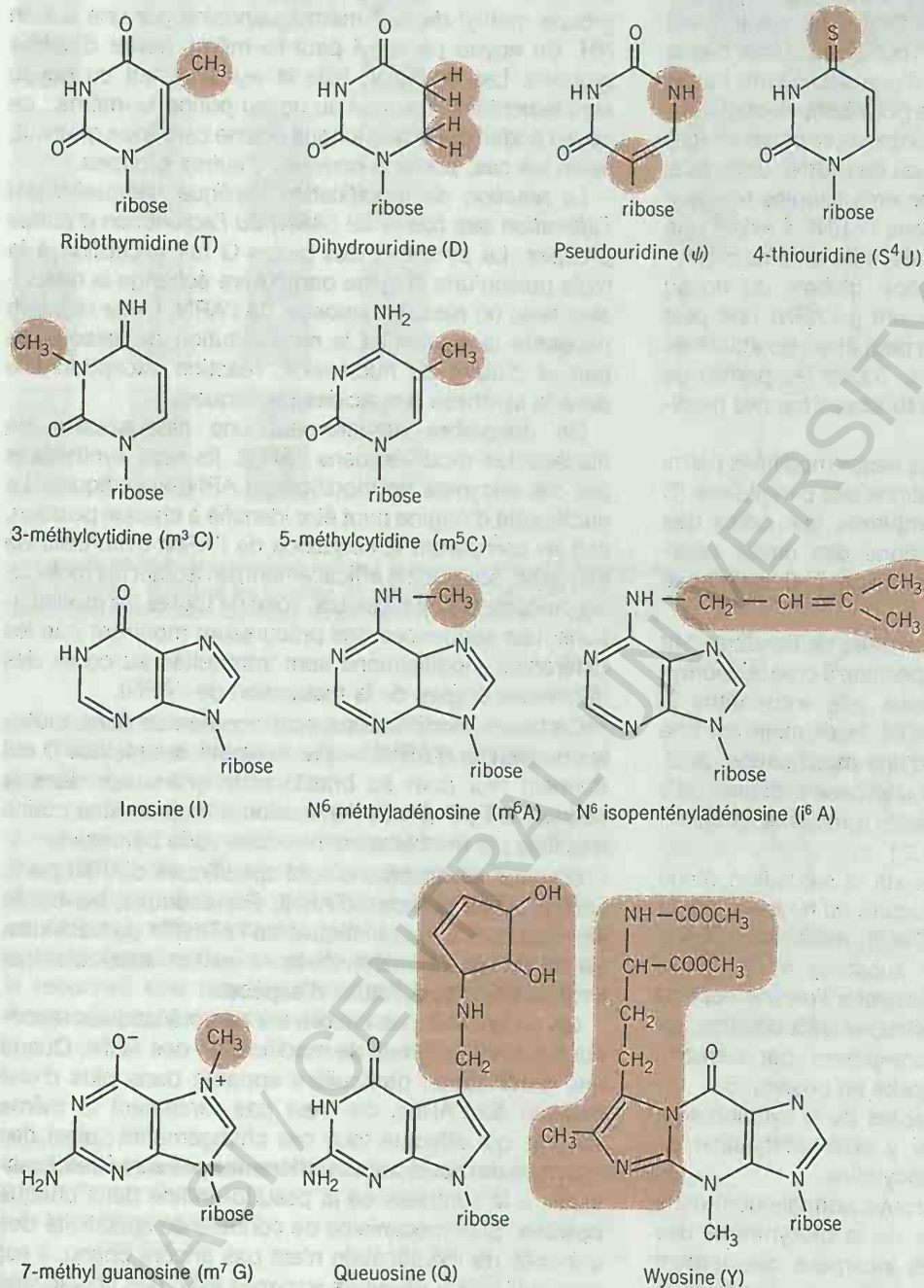


Figure 6.9

Quelques nucléosides modifiés trouvés dans l'ARNt (les modifications apparaissent dans les zones ombrées).

LES MODIFICATIONS DE BASES PEUVENT CONTRÔLER LA RECONNAISSANCE DES CODONS

Vraisemblablement, nous ne comprenons pas encore tous les rôles de la modification. L'effet direct le plus sensible se situe dans l'anticodon lui-même, là où la modification peut influencer sur sa capacité à s'apparier au codon, agissant ainsi sur la signification de l'ARNt. D'autres modifications proches de l'anticodon peuvent aussi agir sur son appariement.

Quand les bases de l'anticodon sont modifiées, d'autres types d'appariement sont réalisables outre ceux répondant aux appariements classiques ou flexibles utilisant A, C, U et G. La **Figure 6.10** montre quelques variétés de liaison codon-anticodon possibles grâce aux combinaisons de modification et de flexibilité.

En fait, certaines combinaisons régulières ne se réalisent jamais car quelques bases sont *toujours* modifiées ; en particulier, U et A ne sont pas utilisées en première position de l'anticodon. D'habitude, dans cette position U apparaît sous une forme modifiée qui peut avoir des propriétés altérées d'appariement. Il semble y avoir une proscription absolue en ce qui concerne l'emploi de A ; habituellement, A est converti en I.

L'inosine, souvent présente en première position de l'anticodon, peut s'apparier à l'une des trois bases U, C et A. Cette aptitude est particulièrement importante pour les codons isoleucine, où AUA code pour l'isoleucine alors que AUG code pour la méthionine. Si nous nous souvenons qu'avec les bases habituelles, il n'est pas possible de reconnaître uniquement A en troisième position, tout ARNt dont l'anticodon débute par U aurait à reconnaître AUG aussi bien que AUA. Ainsi puisque AUA doit être lu comme AUU et AUC, le problème est résolu par l'existence d'un ARNt possédant un anticodon avec I.

Deux autres nucléosides modifiés qui peuvent s'apparier avec trois bases en troisième position du codon sont l'acide uridine-5-oxyacétique et la 5-méthoxyuridine. Ils reconnaissent A et G efficacement mais un peu moins U.

La série de la queuosine et de ses dérivés fournit un autre exemple d'appariements multiples, certains s'effectuant préférentiellement aux autres. Ces bases G modifiées reconnaissent encore C et U mais se lient à U plus fortement.

Une restriction non autorisée par les règles habituelles peut être réalisée par l'utilisation de la 2-thio-uridine dans l'anticodon. Cette modification permet à la base de

continuer à s'apparier avec A, mais prévient sa tolérance pour G.

Toutes ces relations d'appariement ainsi que d'autres mettent bien en valeur le fait qu'il existe plus d'une façon d'établir un jeu d'ARNt capables de reconnaître les 61 codons qui représentent les acides aminés. Dans tout organisme donné, aucun modèle particulier ne prédomine, bien que l'absence d'une voie de modification évite l'emploi de quelques modèles de reconnaissance. Ainsi, une famille particulière de codons peut être lue par des ARNt dont les anticodons diffèrent dans des cellules différentes.

Les ARNt peuvent souvent présenter des réponses qui se chevauchent, un codon particulier peut être lu par plus d'un ARNt. Dans de tels cas, les efficacités de reconnaissances alternatives peuvent être différentes. Et outre la constitution d'un ensemble d'ARNt pouvant reconnaître tous les codons, plusieurs ARNt peuvent aussi répondre aux mêmes codons.

Les prévisions d'appariements flexibles s'accordent très bien aux propriétés observées dans presque tous les ARNt. Néanmoins, il existe des exceptions dans lesquelles les codons reconnus par un ARNt diffèrent de ceux relevant des règles de flexibilité. De tels effets résultent probablement de l'influence des bases avoisinantes et/ou de la conformation de la boucle anticodon au sein de la structure tertiaire globale de l'ARNt. L'importance de la structure de la boucle anticodon est en fait inhérente à l'hypothèse de flexibilité elle-même. Une preuve supplémentaire de l'influence de la structure environnante est fournie par l'isolement de mutants occasionnels chez lesquels le changement d'une base dans une autre région de la molécule altère la capacité de l'anticodon de reconnaître les codons (voir ci-après).

Une autre réaction d'appariement inattendue est fournie par la possibilité pour l'initiateur bactérien, fMet-ARNt^f, de reconnaître à la fois AUG et GUG. Ce comportement inhabituel fait intervenir la troisième base de l'anticodon.

LES MITOCHONDRIES ONT UN NOMBRE RÉDUIT D'ARNt

Des exceptions au code génétique universel apparaissent dans les mitochondries de plusieurs espèces. Le seul changement commun est que dans chaque cas UGA a la même signification que UGG et représente donc le tryptophane au lieu d'une terminaison. Le **Tableau 6.3**

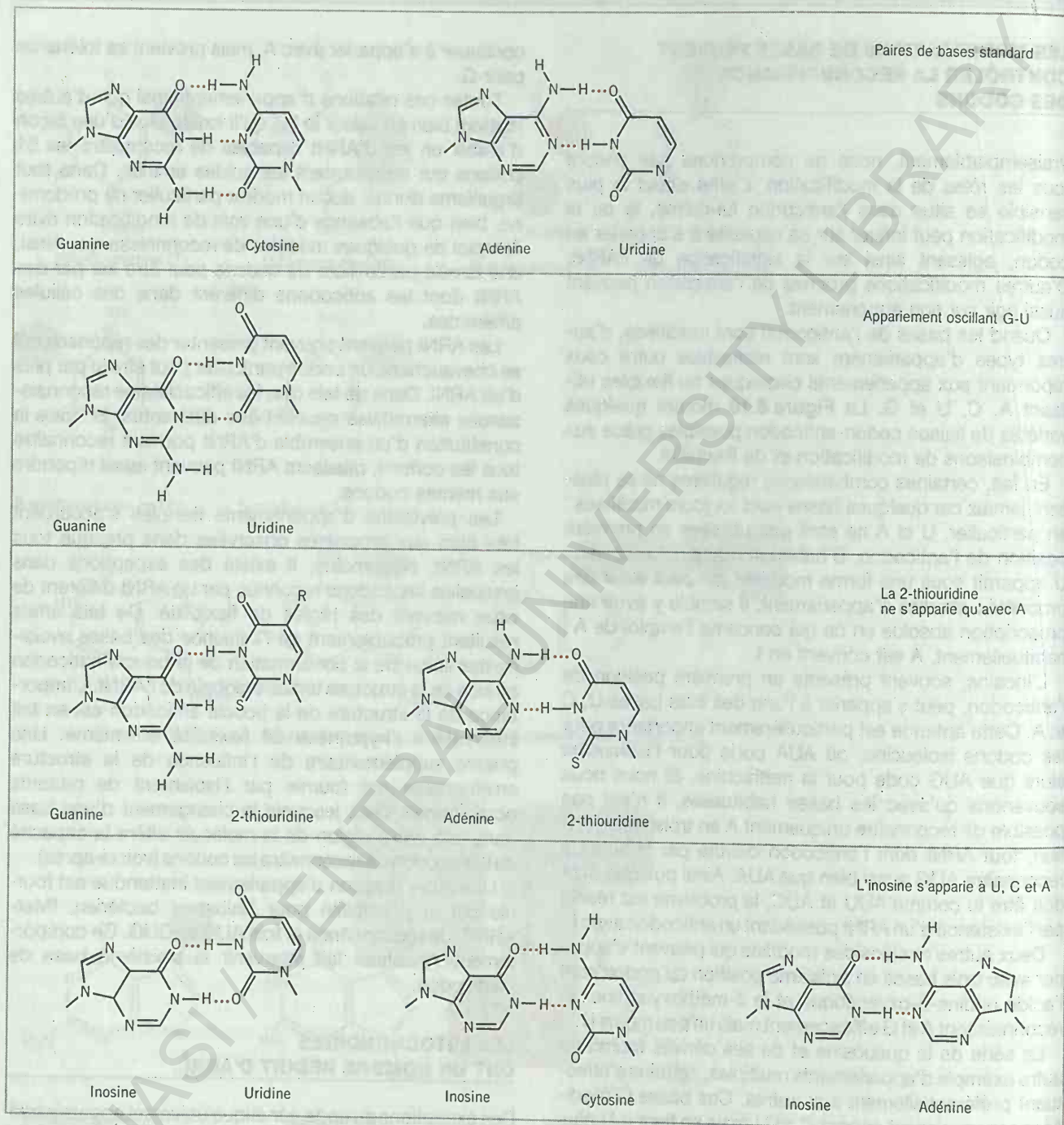


Figure 6.10

La flexibilité dans l'appariement de bases permet à quelques bases en première position de l'anticodon de reconnaître plus d'une base en troisième position du codon.

L'appariement entre les bases standard est élargi des paires G-C et A-U à la paire oscillante G-U. Les modifications de bases peuvent restreindre ou étendre ce modèle. La modification en 2-thiouridine restreint l'appariement à A uniquement car une seule liaison H peut se former avec G. La modification en inosine permet l'appariement avec U, C et A.

Tableau 6.3
Des changements apparaissent dans le code génétique mitochondrial.

Organisme	Codon	Signification dans la mitochondrie	Signification habituelle
Tous	UGA	tryptophane	terminaison
Levure	CUA	thréonine	leucine
Mouche du vinaigre	AGA	sérine	arginine
Mammifère	AG ₆ A	terminaison	arginine
»	AUA	méthionine	isoleucine

résume d'autres changements caractéristiques de chaque espèce.

Certains de ces changements simplifient le code quand deux codons aux significations différentes n'en présentent plus qu'une seule. Ainsi, dans les mitochondries de mammifères, AUA et AUG ne codent plus que pour la méthionine. Un grand nombre de ces variations affecte des codons impliqués soit dans l'initiation, soit dans la terminaison de la synthèse protéique (voir Chapitre 5).

Pourquoi ces changements au sein du code mitochondrial ont-ils pu apparaître ? Puisque les attributions des codons mitochondriaux sont simplifiées (par rapport au code général), ils pourraient représenter un modèle plus primitif. L'autre alternative serait que la simplification aurait évolué en réponse aux besoins particuliers du système de synthèse protéique mitochondrial.

L'existence de changements spécifiques d'espèces implique que la flexibilité mitochondriale a dû être plus importante que la flexibilité nucléaire. Comme la mitochondrie ne synthétise qu'un petit nombre de protéines (environ 10), le problème de rupture due à des changements de sens est beaucoup moins important. Sans doute, les codons altérés n'ont pas été utilisés intensivement dans des régions où les substitutions d'acides aminés auraient été délétères.

Selon l'hypothèse de flexibilité, un minimum de 31 ARNt (l'initiateur non compris) sont nécessaires à la reconnaissance des 61 codons (2 ARNt au moins pour chaque famille de codons et 1 ARNt par paire de codons ou par codon unique). Mais une situation originale existe dans (au moins) les mitochondries de mammifères où n'existent que 23 à 24 ARNt différents. Comment cet ensemble limité d'ARNt s'accommode-t-il de tous les codons ?

La caractéristique principale réside dans la simplification de l'appariement codon-anticodon où un ARNt

reconnaît les quatre membres d'une famille de codons. Ceci réduit à 23 le nombre minimum d'ARNt nécessaires pour répondre à tous les codons internes.

Dans l'ensemble des huit familles de codons, la séquence de l'ARNt contient une base U non modifiée en première position de l'anticodon. Les codons restants se regroupent par paires dans lesquelles tous les codons se terminant par une pyrimidine sont lus par G dans l'anticodon et tous les codons se terminant par une purine sont lus par U dans l'anticodon, comme le prévoit l'hypothèse de flexibilité. La complication entraînée par le codon unique UUG est évitée par le changement dans le code permettant de lire UGA avec UGG en tant que tryptophane ; et chez les mammifères, AUA cesse de représenter l'isoleucine pour être lu avec AUG comme une méthionine. Ceci permet à tous les codons ne faisant pas partie de familles d'être lus en 14 paires.

Les 23 gènes d'ARNt codent pour : un initiateur, 14 ARNt représentant des paires et 8 ARNt représentant des familles. Ainsi, les deux codons non-sens habituels UAG et UAA ne sont reconnus par aucun ARNt, de même qu'une paire de codons arginine (AG ₆) qui, elle aussi peut servir à la terminaison (un autre changement dans le code). Des règles identiques sont suivies par les mitochondries de champignons.

L'utilisation de U dans les anticodons qui répondent à la fois à l'ensemble des familles de codons et aux paires qui se terminent par une purine, est paradoxale. La difficulté semble être évitée par la modification de U dans les ARNt qui ne répondent qu'aux paires de codons. Dans l'anticodon, U non modifiée doit pouvoir s'apparier de façon flexible avec U et C aussi bien qu'avec A et G comme le montre la Figure 6.10. S'il s'agit d'une capacité inhérente, elle peut expliquer pourquoi U est toujours modifiée en première position des anticodons de bactéries et du cytoplasme eucaryote (bien que nous nous demandions pourquoi U n'est jamais utilisée pour y reconnaître des familles entières de codons).

LES ARNt MUTÉS LISENT DIFFÉRENTS CODONS

L'isolement d'ARNt mutés a constitué l'un des outils les plus puissants pour analyser la capacité d'un ARNt à répondre à son ou ses codons, et à déterminer les effets que les différentes parties de la molécule d'ARNt peuvent avoir sur la reconnaissance codon-anticodon.

Les ARNt mutés sont isolés grâce à leur pouvoir d'outrepasser les effets de mutations sur des gènes qui codent pour des protéines. Nous avons déjà décrit la

terminologie selon laquelle une mutation qui est capable d'annuler les effets d'une autre mutation est appelée **mutation de type supresseur** (voir Chapitre 2).

Dans les systèmes ARNt supresseurs, la mutation primaire change un codon dans l'ARNt si bien que la protéine synthétisée n'est plus fonctionnelle ; la mutation secondaire, suppressive, en changeant l'anticodon de l'ARNt lui permet de reconnaître le codon muté au lieu de (ou aussi bien que) son codon cible d'origine. L'acide aminé qui est alors incorporé restaure la fonction de la protéine. Les supresseurs sont appelés **non-sens** ou **faux-sens**, relativement à la mutation d'origine.

Dans une cellule sauvage, une mutation non-sens n'est reconnue que par un facteur de libération, mettant fin à la synthèse de la protéine. La mutation suppressive crée un aminoacyl-ARNt qui peut reconnaître le codon de terminaison ; en insérant un acide aminé, elle permet à la synthèse protéique de progresser au-delà du site de la mutation non-sens.

Cette nouvelle capacité du système de traduction rend possible la synthèse de la totalité de la chaîne polypeptidique comme l'illustre la **Figure 6.11**. Si l'acide aminé incorporé par suppression est différent de l'acide aminé original présent dans ce site de la protéine sauvage, l'activité de cette dernière peut être réduite.

Les supresseurs non-sens se répartissent en trois classes, une pour chaque type de codon de terminaison. Le **Tableau 6.4** regroupe les propriétés de quelques supresseurs les mieux caractérisés.

Les supresseurs ambre ont été les plus faciles à caractériser. Chez *E. coli*, 6 ARNt au moins ont été mutés pour reconnaître les codons UAG. Tous les ARNt supresseurs ambre ont l'anticodon $\bar{CUA}$, dérivé dans chaque cas du type sauvage par un seul changement de base. Le site de mutation peut atteindre chacune des trois bases de l'anticodon, ainsi pour *supD*, *supE* et *supF*. Chaque ARNt supresseur ne reconnaît *que* le codon AUG au lieu de son ou ses codons précédents. Les acides aminés incorporés sont la sérine, la glutamine ou la tyrosine, identiques à ceux portés par les ARNt sauvages correspondants.

Les supresseurs ocre proviennent aussi de mutations dans l'anticodon. Les mieux connus sont *supC* et *supG*, qui incorporent la tyrosine et la lysine en réponse aux *deux codons* : ocre (UAA) et ambre (UAG). Ceci est exactement conforme à l'hypothèse de flexibilité selon laquelle UAA ne peut pas être reconnu seul.

Les supresseurs de UGA sont intéressants. Un type (*supU*) est un mutant de l'ARNt du tryptophane. L'anticodon de l'ARNt^{Trp} sauvage est $\bar{CCA}$ qui ne reconnaît que

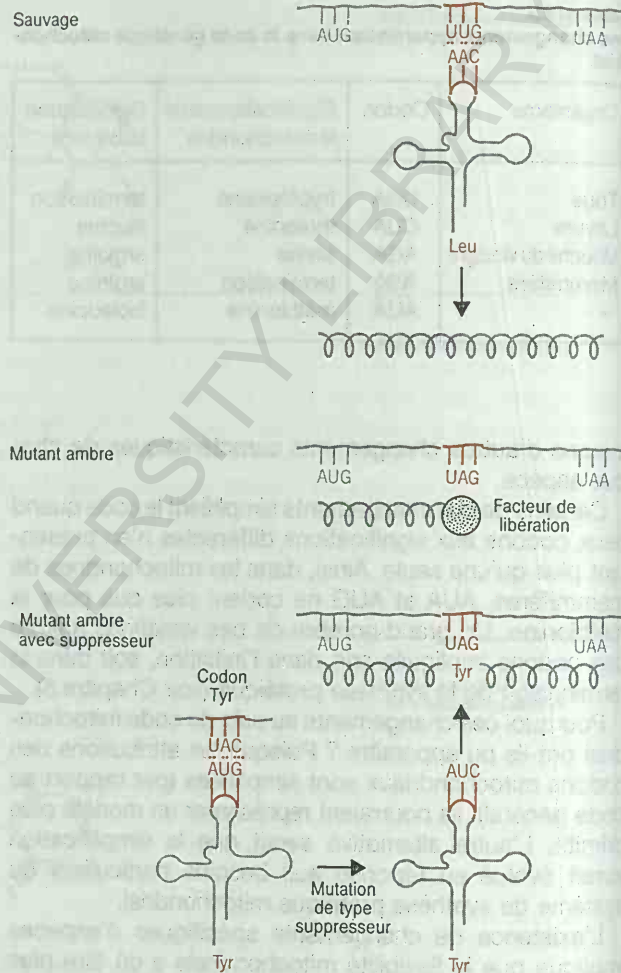


Figure 6.11

Les mutations non-sens peuvent être supprimées par un ARNt dont l'anticodon est muté.

Le gène sauvage contient un codon UUG spécifique de la leucine, reconnu par un ARNt dont l'anticodon est $\bar{CAA}$. Par mutation, le codon devient UAG, il est alors reconnu par un facteur de libération. A la suite d'une seconde mutation, l'anticodon de l'ARNt tyrosine passe de $\bar{GUA}$ à $\bar{CUA}$, étant alors reconnu par le codon ambre. La suppression entraîne la synthèse de la totalité de la protéine dans laquelle le résidu Leu d'origine a été remplacé par Tyr.

le codon UGG. L'ARNt supresseur possède l'anticodon muté $\bar{UCA}$ qui reconnaît à la fois son codon d'origine et le codon de terminaison UGA.

Un autre supresseur UGA possède une propriété totalement inattendue. Il est dérivé du même ARNt^{Trp}, mais son unique mutation résulte de la substitution de G

Tableau 6.4

Les ARNt non-sens supprimeurs sont engendrés par des mutations dans l'anticodon.

Locus	ARNt	Type sauvage		Supprimeur	
		Codons reconnus	Anti-codon	Anti-codon	Codons reconnus
<i>supD</i> (<i>su1</i>)	Ser	UCG	CGA	CUA	UAG
<i>supE</i> (<i>su2</i>)	Gln	CAG	CUG	CUA	UAG
<i>supF</i> (<i>su3</i>)	Tyr	UA ^G _G	GUA	CUA	UAG
<i>supC</i> (<i>su4</i>)	Tyr	UA ^G _G	GUA	UUA	UAA _G
<i>supG</i> (<i>su5</i>)	Lys	AA ^A _G	UUU	UUA	UAA _G
<i>supU</i> (<i>su7</i>)	Trp	UGG	CCA	UCA	UGA _G

La base qui a changé dans l'anticodon muté est indiquée en caractère gras.

Nomenclature. A l'origine, les loci étaient connus en tant que *su1*, *su2*, etc., comme cela apparaît, entre parenthèses, avec *su** indiquant la forme mutée de type supprimeur du gène et *su* - la forme de type sauvage. Actuellement, les loci sont redésignés par *sup* suivi d'une lettre allant de A à V. Tous les loci décrits ici sont des gènes structuraux qui codent pour des ARNt.

par A en position 24. Ce changement remplace une paire G-U dans le bras D par une paire A-U, augmentant la stabilité de l'hélice. La séquence de l'anticodon reste la même que celle du type sauvage : CCA. Donc la mutation dans le bras D doit d'une certaine façon altérer la conformation de la boucle anticodon, autorisant un appariement CCA/UGA, au cours d'un appariement flexible inhabituel entre C et A. L'ARNt supprimeur continue à reconnaître son codon usuel UGG.

Une réponse du même ordre existe avec un ARNt eucaryote. Le foie de bovin contient un ARNt^{Ser} avec un anticodon ^mCCA. Selon l'hypothèse de flexibilité, cet ARNt devrait répondre au codon tryptophane UGG ; mais en fait, il répond au codon de terminaison UGA. Il est donc possible que dans cette situation, UGA soit supprimé naturellement.

L'importance de ces observations repose sur la démonstration que la reconnaissance codon-anticodon de l'ARNt sauvage ou muté ne peut être déduite totalement des séquences de triplets concernés, mais peut être influencée par d'autres caractéristiques de la molécule.

Tous ces supprimeurs non-sens sont dus à des mutations ponctuelles dans le gène de structure qui code pour l'ARNt. Mais en principe, les ARNt supprimeurs pourraient aussi être la conséquence de la modification des bases dans l'anticodon. De tels changements résulteraient de la mutation du gène qui code pour une enzyme modifiant l'ARNt.

Les mutations **faux-sens** transforment un codon qui représente un acide aminé en un codon qui code pour un autre acide aminé, ce dernier ne pouvant fonctionner dans la protéine à la place du résidu d'origine. (En théorie, toute substitution d'acides aminés constitue une mutation faux-sens, mais en pratique, elle n'est détectée que si elle fait varier l'activité de la protéine.) La mutation peut être supprimée par l'insertion soit de l'acide aminé d'origine, soit d'un autre acide aminé acceptable par la protéine.

La **Figure 6.12** démontre que la mutation faux-sens peut s'accomplir de la même façon que la mutation non-sens, par la mutation de l'anticodon d'un ARNt portant un acide aminé acceptable de telle sorte qu'il réponde au codon muté. Ainsi, la mutation faux-sens

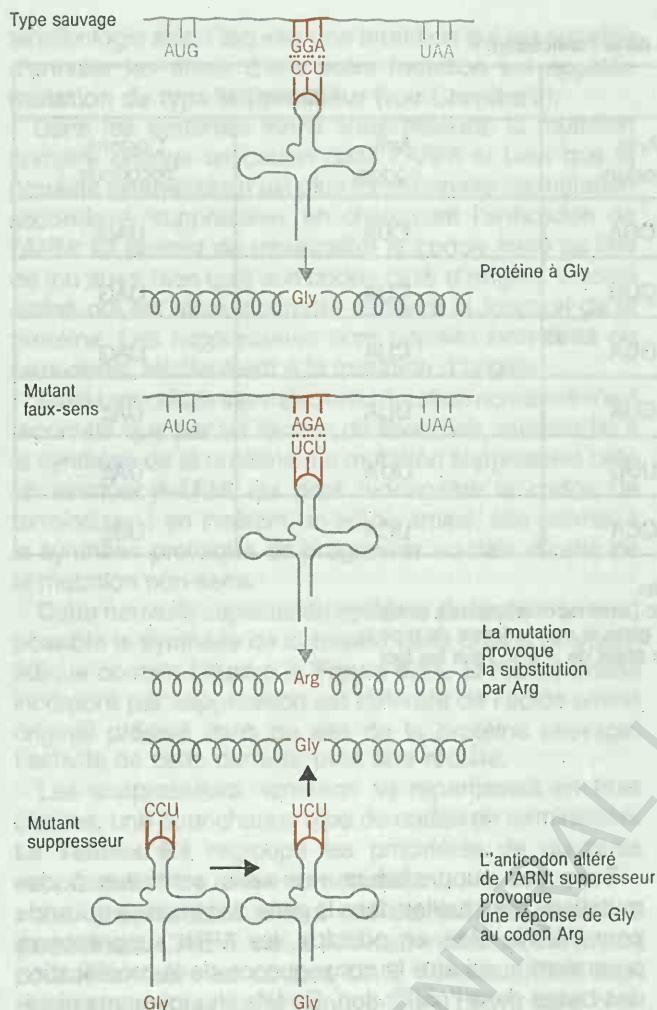


Figure 6.12

La mutation faux sens apparaît quand l'anticodon de l'ARNt est muté et qu'il répond alors au « faux » codon.

La mutation d'un codon GGA en AGA provoque le remplacement dans la protéine mutée de la Gly par Arg. Si l'anticodon de l'ARNt^{Gly} est muté de UCC en UCU, il insère Gly en réponse au codon AGA qui représenterait Arg. Il est à noter que l'ARNt^{Arg} sauvage et l'ARNt^{Gly} supprimeur répondront tous les deux à AGA, si bien que la suppression n'est que partielle.

implique le changement de signification du codon d'un acide aminé en un autre acide aminé.

Une autre façon de créer des supprimeurs faux-sens est la mutation dans l'ARNt (ou synthétase) qui provoque

le changement de l'acide aminé porté par l'ARNt. Bien que l'ARNt continue à répondre au(x) même(s) codon(s), l'insertion d'un acide aminé différent pourrait entraîner la suppression. Ce type de suppression est plutôt inhabituel, probablement du fait qu'il nécessite une mutation qui crée un nouveau site de reconnaissance.

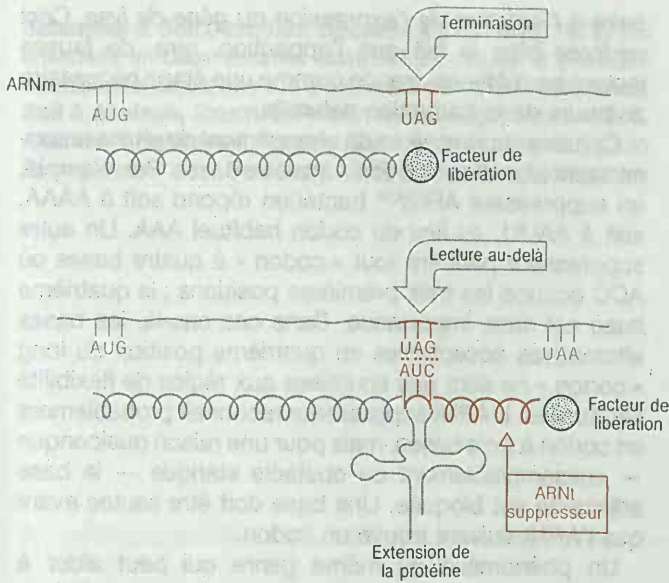
La plupart des supprimeurs ont été caractérisés chez *E. coli* bien que certains mutants comparables aient été obtenus chez d'autres bactéries (en particulier *S. typhimurium*). Les connaissances relatives à l'apparition et à la suppression éventuelle de mutants non-sens et faux-sens chez les eucaryotes sont beaucoup moins étendues, mais le tableau est probablement identique pour les codons ocre et ambre, malgré les hypothèses selon lesquelles UAG pourrait quelquefois n'être que partiellement supprimé.

LES ARNt SUPPESSEURS SONT EN COMPÉTITION VIS-A-VIS DE LEURS CODONS

Il existe une différence intéressante entre le mode de reconnaissance habituel d'un codon par son propre aminoacyl-ARNt et la situation dans laquelle une mutation permet à un ARNt supprimeur de reconnaître un nouveau codon. Dans la cellule sauvage, un codon donné n'a qu'une seule signification, soit un acide aminé, soit un signal de terminaison. Mais dans une cellule portant une mutation de type supprimeur, deux alternatives se présentent au codon muté : être reconnu par l'ARNt supprimeur ou être lu avec sa signification habituelle.

Un ARNt supprimeur non-sens peut entrer en compétition avec les facteurs de libération qui reconnaissent son ou ses codons de terminaison cible. Un ARNt supprimeur faux-sens peut entrer en compétition avec les ARNt qui répondent correctement à son nouveau codon. L'étendue de la compétition influencera la valeur de la suppression ; ainsi l'efficacité d'un supprimeur particulier peut dépendre non seulement de l'affinité entre son anticodon et le codon cible, mais aussi de sa concentration dans la cellule, et des paramètres gouvernant la compétition entre les réactions de terminaison ou d'insertion.

L'efficacité de lecture d'un codon, quel qu'il soit, peut être influencée par sa localisation. Ainsi, l'efficacité d'une suppression non-sens par un ARNt donné peut varier largement, en fonction du contexte du codon. On ne comprend pas encore l'effet des bases avoisinantes sur l'ARNm sur la reconnaissance codon-anticodon, mais

**Figure 6.13**

Les suppresseurs non-sens lisent aussi au-delà des codons de terminaison naturels, synthétisant des protéines plus longues que les protéines sauvages.

elles peuvent changer la fréquence avec laquelle un codon est reconnu par un ARNt particulier de plus de 10 fois. La base en 3' du codon semble avoir un effet particulièrement fort.

Un suppresseur non-sens est isolé du fait de sa capacité à répondre à un codon non-sens muté. Mais la même séquence triplet constitue un des signaux normaux de terminaison dans la cellule ! L'ARNt muté qui supprime la mutation non-sens doit en principe pouvoir supprimer la terminaison normale à la fin de tout gène qui utilise ce codon. La **Figure 6.13** montre que ce **non respect du codon stop** (readthrough) provoque la synthèse d'une protéine plus longue avec un matériel C-terminal supplémentaire. La protéine allongée se terminera à la séquence triplet de terminaison suivante trouvée dans le cadre de lecture. Toute suppression étendue de terminaison est sans doute délétère pour la cellule car elle produit des protéines plus longues dont les fonctions sont par là-même altérées.

Les suppresseurs ambre ont tendance à être relativement efficaces, atteignant des taux de 10 à 50 %, selon

le système. Si nous supposons que la cellule ne peut pas tolérer un tel niveau de non respect du codon stop à la fin des protéines naturelles, ceci indique que l'utilisation des codons ambre en fin de gène est moins fréquente que celle des autres codons de terminaison.

Les suppresseurs ocre sont difficiles à isoler. Ils sont toujours moins efficaces ; habituellement leurs activités sont inférieures à 10 %. Tous les suppresseurs ocre poussent lentement, ce qui prouve que la double suppression d'UAA et UAG est préjudiciable à *E. coli*, sans doute du fait que le codon ocre est utilisé beaucoup plus fréquemment comme un signal de terminaison habituel.

De la même façon un suppresseur faux-sens génique est probablement l'agent de mutation d'un autre gène. Un suppresseur corrige une mutation en substituant un acide aminé à un autre au site muté. Mais dans d'autres localisations, la même substitution remplacera l'acide aminé sauvage par un nouvel acide aminé. Le changement peut inhiber la fonction normale de la protéine.

La cellule a donc un dilemme à résoudre : elle doit supprimer un codon muté en une localisation tout en s'abstenant de faire trop varier sa signification normale en d'autres localisations. L'absence de tout suppresseur faux-sens fort s'explique donc par les effets délétères qu'engendrerait une substitution d'acides aminés générale et efficace.

Une mutation qui crée un ARNt suppresseur peut avoir deux conséquences. Premièrement, elle permet à l'ARNt de reconnaître un nouveau codon. Deuxièmement, elle ne *permet plus*, quelquefois, à l'ARNt de reconnaître les codons auxquels il répondait précédemment. Il est significatif que tous les suppresseurs ambre hautement efficaces dérivent de la mutation d'une copie d'un ensemble d'ARNt redondant. Dans ces situations, la cellule possède plusieurs ARNt qui peuvent répondre au codon reconnu à l'origine par l'ARNt sauvage. La mutation n'abolit donc pas la reconnaissance des anciens codons qui continuent à être utilisés de façon adéquate par les ARNt de l'ensemble.

En contraste avec cette situation, un autre ARNt suppresseur ambre est **létal récessif**. Dérivé de *supU*, le seul gène qui code pour le tryptophane ARNt chez *E. coli*, il peut agir comme un suppresseur quand la cellule possède deux copies du gène, une sauvage et l'autre mutée. Mais si l'unique copie est mutée, la cellule meurt.

La mutation transforme l'anticodon de l'ARNt de $\overline{CCA}$ en $\overline{CUA}$. Mais cette altération laisse la cellule dépourvue d'un ARNt qui puisse répondre à UGG, changement qui s'avère létal.

L'ARNt PEUT INFLUER SUR LE CADRE DE LECTURE

Le cadre de lecture d'un messenger est habituellement invariable. La traduction commence au codon AUG et continue le long des triplets jusqu'au codon de terminaison. La lecture ne tient pas compte du sens : l'insertion ou la délétion d'une base provoque une mutation de changement de phase (frame shift) dans laquelle le cadre de lecture n'est plus le même au-delà du site de mutation. Ribosomes et ARNt avancent inéluctablement au long des triplets en synthétisant une série d'acides aminés totalement différente.

La suppression du changement de phase restaure le cadre de lecture original. Ce mécanisme de compensation de délétion ou d'insertion de bases a déjà été discuté (voir Chapitre 2). Cependant des supprimeurs de changement de phase extragéniques ont aussi été trouvés.

Un type de supprimeur externe de changement de phase corrige le cadre de lecture quand une mutation résulte de l'insertion d'une base supplémentaire dans une extension de résidus identifiés. Par exemple, un G peut avoir été inséré dans une série de plusieurs bases G contiguës. Le supprimeur de changement de phase est un ARNt^{Gly} qui possède une base additive insérée dans la boucle de l'anticodon, transformant sa séquence de triplets habituelle CCC en une séquence quadruplet CCCC. L'ARNt supprimeur reconnaît un « codon » à quatre bases.

Un type proche de suppression de changement de phase a été trouvé dans la mitochondrie de levure, où les mutations qui insèrent ou délètent un seul résidu T dans une série de cinq résidus T sont toutes perdues. La raison semble en être que l'ARNt^{Phe} qui répond au codon UUU permet quelquefois au ribosome de glisser d'une base dans un sens ou un autre supprimant ainsi la mutation. Ce glissement semble être un phénomène normal quoique rare (fréquence inférieure à 5 %).

L'ARN du phage MS2 est un autre exemple de situation où un changement de phase occasionnel est un événement normal. Un changement de phase (dont le mécanisme est inconnu) entraîne la reconnaissance par le ribosome d'un codon de terminaison prématuré dans le nouveau cadre de lecture. Le ribosome terminant la traduction peut ensuite reconnaître le codon d'initiation du gène de lyse qui n'en est distant que de quelques bases.

En l'absence d'un changement de phase, le ribosome ne termine pas la traduction et ne reconnaît pas le codon d'initiation du gène de lyse. *Donc la terminaison dépendante du changement de phase est un préalable néces-*

saire à l'initiation de l'expression du gène de lyse. Ceci renforce bien le fait que l'apparition, rare, de fausse lecture peut être envisagée comme une étape nécessaire au cours de la traduction naturelle.

Certains supprimeurs de changement de phase reconnaissent plus d'un « codon » à quatre bases. Par exemple, un supprimeur ARNt^{Lys} bactérien répond soit à AAAA, soit à AAUU, au lieu du codon habituel AAA. Un autre supprimeur peut lire tout « codon » à quatre bases où ACC occupe les trois premières positions ; la quatrième base est sans importance. Dans ces cas-là, les bases alternatives acceptables en quatrième position du long « codon » ne sont pas soumises aux règles de flexibilité habituelles. L'ARNt supprimeur reconnaît probablement un codon à trois bases, mais pour une raison quelconque — vraisemblablement un obstacle stérique — la base adjacente est bloquée. Une base doit être sautée avant que l'ARNt suivant trouve un codon.

Un phénomène du même genre qui peut aider à comprendre la base de cette action est la possibilité de provoquer un changement de phase en manipulant des systèmes *in vitro* qui ne contiennent que des constituants normaux. Un large excès ou un déficit en un aminoacyl-ARNt particulier peut produire des changements de phase. Le temps de latence de réponse au codon à trois bases au site A permet sans doute l'accès d'un ARNt qui reconnaît un codon chevauchant (hors phase).

Ceci suggère un modèle uniforme de contrôle de la translocation ribosomale. En avançant, le ribosome déplace l'ARNt qui était au site A afin qu'il occupe correctement le site P. Quand un ARNt est lié à un « codon » à quatre bases, ou bien s'il s'étend stériquement au-delà de son codon à trois bases habituel, l'aminoacyl-ARNt suivant se lie à un triplet hors phase dans le site A. Mais en cours de translocation, le ribosome le dépose dans le site P, de façon telle que les aminoacyl-ARNt qui suivent rencontrent un codon triplet au site A comme d'habitude. Ainsi la géométrie de l'interaction ARNt/ARNm est utilisée pour déterminer la distance de déplacement du ribosome.

POUR EN SAVOIR PLUS

Les données concernant l'ARNt ont fait l'objet de nombreux ouvrages. Dans Altman's (Ed.), *Transfer RNA* (MIT Press, Cambridge, 1978), Clark a discuté la structure primaire et secondaire (pp. 14-17) et Kim la structure tertiaire (pp. 248-293). Les bases modifiées ont été étudiées par Nishimura (pp. 168-195). Igloi & Cramer expriment leur point de vue sur l'aminoacyl-ARNt synthétase (pp. 294-340) de même que

Schimmel & Soll (*Ann. Rev. Biochem.* 48, 601-648, 1979). Un ensemble en deux volumes composé de revues et d'articles scientifiques donne une vue d'ensemble de l'ARNt : Schimmel, Soll & Abelson, *Structure, Properties and Recognition*, et Soll, Abelson & Schimmel, *Transfer RNA : Biological Aspects* (Cold Spring Harbor Lab., New York, 1979). Un travail original sur les

propriétés codantes de l'ARNt et ses variations via les mutations a été traité dans le Chapitre 5 de Lewin, *Gene Expression*, 1, *Bacterial Genomes* (Wiley, New York, 1974). La suppression a été revue par Korner, Feinstein & Altman (pp. 105-135 dans le livre d'Altman), et des développements plus récents ont été résumés par Murgola (*Ann. Rev. Genet.* 19, 57-80, 1985).

CHAPITRE 7

LE RIBOSOME : USINE DE LA TRADUCTION

Le ribosome se comporte comme une petite usine itinérante qui se déplace le long de la matrice, engagée dans des cycles rapides de synthèse de liaison peptidique. Les aminoacyl-ARNt entrent et sortent de la particule à un rythme effrayant, déposant leurs acides aminés, les facteurs d'élongation s'associant ou se dissociant de façon cyclique. Avec ses facteurs accessoires, le ribosome fournit la gamme complète des activités nécessaires à toutes les étapes de la synthèse protéique.

Les ribosomes sont un constituant cellulaire majeur. Dans une bactérie en phase de croissance active, il peut y avoir jusqu'à 20 000 ribosomes par génome (voir Tableau 5.1). Ils contiennent environ 10 % des protéines bactériennes totales et représentent 80 % de la masse totale de l'ARN cellulaire. Dans les cellules eucaryotes, la proportion des protéines totales qu'ils représentent est moindre mais ils sont plus nombreux en valeur absolue et néanmoins contiennent la presque totalité de la masse d'ARN cellulaire. Le nombre de ribosomes (chez les procaryotes comme chez les eucaryotes) est en relation directe avec l'activité de synthèse protéique de la cellule.

Les ribosomes bactériens sont liés aux ARNm eux-mêmes encore reliés à l'ADN. Dans le cytoplasme eucaryotique, les ribosomes sont habituellement associés au cytosquelette, une matrice fibreuse. Dans certaines cellu-

les eucaryotes, une partie des ribosomes est associée aux membranes du réticulum endoplasmique. La caractéristique commune est que les ribosomes engagés dans la traduction ne sont pas à l'état libre dans la cellule mais sont associés, directement ou indirectement, aux structures cellulaires.

Tous les ribosomes peuvent être dissociés en deux sous-unités, dont l'une est environ deux fois plus grosse que l'autre. Chaque sous-unité contient un ARNr majeur et un très grand nombre de protéines différentes, presque toutes présentes sous forme d'une seule copie par sous-unité. La sous-unité la plus grande peut contenir outre l'ARNr majeur une ou des molécules d'ARN plus petites. Le fait que la séquence de l'ARNr soit presque toujours invariable dans une cellule, bien que codée par plusieurs gènes, met bien en évidence l'importance de l'ARNr. Cette conservation plaide en faveur d'une sélection vis-à-vis des variations de sa séquence.

Un rôle majeur des ARNr est nettement structural. Les protéines se lient à des sites particuliers de l'ARNr, selon un ordre spécifique nécessaire à l'assemblage de chaque sous-unité. Cependant, l'intérêt des ribosomes ne réside pas uniquement dans leur fonction mais dans le procédé au cours duquel ils s'auto-assemblent à partir de l'ARN et des protéines.

Le ribosome possède plusieurs centres actifs, chacun constitué d'un groupe particulier de protéines. Certaines fonctions catalytiques peuvent être attribuées à des protéines particulières mais aucune de ces activités ne peut être réalisée par ces protéines ou groupes de protéines isolées. Le ribosome représente un ensemble d'enzymes, chacune étant active uniquement au sein de la structure globale appropriée, et dont toutes les activités coordonnées aboutissent au mécanisme de la traduction. Beaucoup de protéines (ainsi que l'ARNr) peuvent intervenir principalement dans l'élaboration de la structure générale qui permet aux différents sites actifs d'entrer en relation ; elles ne participent pas nécessairement de façon directe aux réactions de synthèse.

LES RIBOSOMES SONT DES PARTICULES RIBONUCLÉOPROTÉIQUES COMPACTES

La construction des ribosomes peut se réaliser de différentes façons. Les ribosomes de bactéries, de cytoplasme eucaryote et d'organelles présentent des variations notables dans leur taille globale et les proportions respectives d'ARN et de protéines. Le **Tableau 7.1** regroupe les constituants du ribosome bactérien en prenant pour exemple *E. coli*.

Tous les ribosomes d'une bactérie sont identiques.

Chez *E. coli*, les ARN ribosomaux ont été séquencés et la séquence en acides aminés de la plupart des protéines a été déterminée. La petite sous-unité (30S) est constituée de l'ARNr 16S et de 21 protéines. La grande sous-unité (50S) contient l'ARNr 23S, le petit ARNr 5S et 31 protéines. Exception faite d'une protéine présente en quatre copies par ribosome, chaque protéine n'est représentée que par une copie. L'ARNr constitue la plus grande partie de la masse de la particule.

Les ribosomes cytoplasmiques des eucaryotes supérieurs sont plus grands que les ribosomes bactériens. Le **Tableau 7.2** résume un exemple chez le mammifère. Le contenu total en ARN et en protéines est plus grand ; les molécules d'ARN majeur sont plus longues et un plus grand nombre de protéines sont présentes. Ces dernières sont probablement toutes ou presque toutes présentes en quantités stoechiométriques. L'ARN est néanmoins toujours le constituant prédominant en masse.

Les ribosomes prennent des formes diverses dans les organelles. Quelques exemples sont réunis dans le **Tableau 7.3**. Leur taille peut varier de celle des ribosomes cytoplasmiques jusqu'à une taille inférieure à celle des ribosomes bactériens.

Les formes des ribosomes de bactéries, de chloroplastes et du cytoplasme eucaryote sont en général similaires (les ribosomes mitochondriaux n'ont pas été bien caractérisés). La **Figure 7.1** est une représentation schématique de la petite sous-unité bactérienne dont la forme est

Tableau 7.1
Le ribosome d'*E. coli* contient trois molécules d'ARN et 52 protéines.

	Ribosome	Petite sous-unité	Grande sous-unité
Vitesse de sédimentation	70S	30S	50S
Masse (daltons)	2 520 000	930 000	1 590 000
ARN majeurs		16S = 1 541 bases	23S = 2 904 bases
ARN mineurs			5S = 120 bases
Masse d'ARN	1 664 000	560 000	1 104 000
Proportion d'ARN	66 %	60 %	70 %
Nombre de protéines		21 polypeptides	31 polypeptides
Masse des protéines	857 000	370 000	487 000
Proportion des protéines	34 %	40 %	30 %

Les protéines de la petite sous-unité sont numérotées S1-S21. Les protéines de la grande sous-unité sont numérotées L1-L34, en raison de quelques appellations initiales inappropriées.

Tableau 7.2

Les ribosomes cytoplasmiques de mammifère (foie de rat) contiennent quatre molécules d'ARN et 82 protéines.

	Ribosome	Petite sous-unité	Grande sous-unité
Vitesse de sédimentation	80S	40S	60S
Masse (daltons)	4 220 000	1 400 000	2 820 000
ARN majeurs		18S = 1 874 bases	28S = 4 718 bases
ARN mineurs			5,8S = 160 bases 5S = 120 bases
Masse d'ARN	2 520 000	700 000	1 820 000
Proportion d'ARN	60 %	50 %	65 %
Nombre de protéines		33 polypeptides	49 polypeptides
Masse des protéines	1 700 000	700 000	1 000 000
Proportion des protéines	40 %	50 %	35 %

quelque peu aplatie. La **Figure 7.2** montre que la grande sous-unité a une structure plus sphérique.

La configuration asymétrique du ribosome complet 70S est illustrée par la **Figure 7.3**. Le resserrement entre le corps et la tête de la petite sous-unité est aligné avec l'encoche de la grande sous-unité, si bien que la plateforme de la petite sous-unité s'adapte dans la grande sous-unité. Il peut subsister un espace ou un « tunnel » entre les deux sous-unités.

Quelques photographies en microscopie électronique de sous-unités et de ribosomes bactériens complets sont présentées **Figure 7.4**, ainsi que leurs modèles, dans la même orientation.

Bien qu'ils soient plus grands, les ribosomes cytoplasmiques de mammifères ont une structure apparente identique. Il existe une similitude suggestive entre les structures des petites sous-unités bactériennes et eucaryotes ; les grandes sous-unités semblent plus divergentes. Bien qu'un modèle détaillé n'ait pas encore été

élaboré, la relation entre les sous-unités de mammifères ressemble à celle des ribosomes procaryotes.

Une grande attention a été portée à la localisation de l'ARNm dans le ribosome. Une idée généralement répandue est qu'il vient s'ajuster entre les sous-unités ou près de leur jonction. Quelques modèles représentent l'ARNm comme enfilé à travers le ribosome, mais sa localisation peut aussi bien être superficielle. La **Figure 7.5** montre de façon évidente que les deux ARNt sont relativement grands par rapport au ribosome ; il semble plus que probable qu'ils s'insèrent dans de grandes « fentes » s'ouvrant en surface plutôt que complètement à l'intérieur de cavités internes.

Tableau 7.3

Variation de la taille des ribosomes dans les organelles.

Organelle	Taille du ribosome	Proportion d'ARN
Chloroplaste d' <i>Euglena</i>	70S	~ 70 %
Mitochondrie de maïs	78S	~ 70 %
Mitochondrie de levure	73S	~ 70 %
Mitochondrie de <i>Xenopus</i>	60S	< 30 %
Mitochondrie humaine	60S	< 30 %

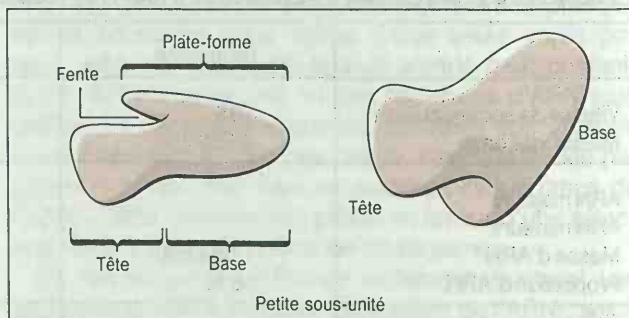
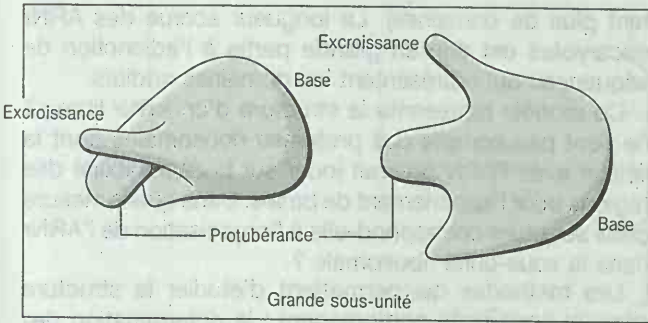


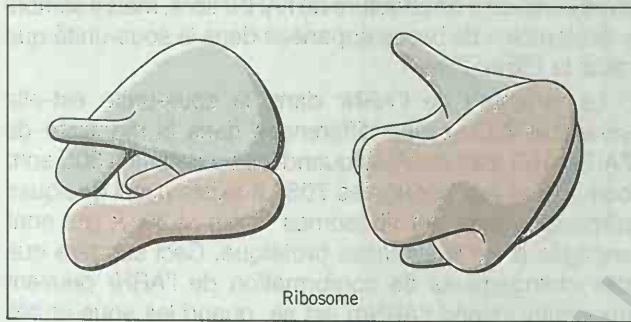
Figure 7.1

La sous-unité 30S possède une forme allongée et asymétrique, d'environ $55 \times 220 \times 220 \text{ \AA}$, avec une région étranglée et une « fente » qui sépare la « tête » de la « base ».

Il existe un angle de 90° entre les deux images qui représentent la sous-unité.

**Figure 7.2**

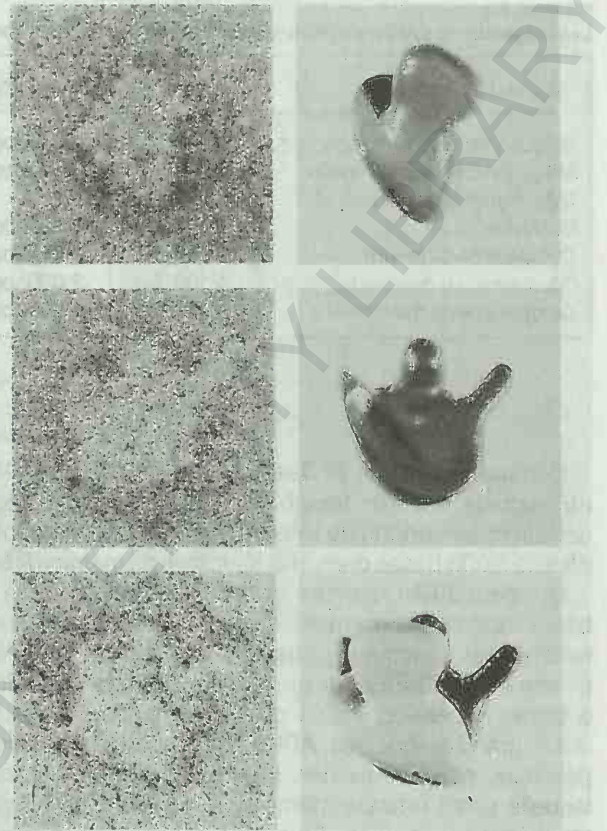
Les deux images montrent que la sous-unité 50S a une structure plutôt compacte, environ $150 \times 200 \times 200 \text{ \AA}$, à partir de laquelle une « protubérance centrale » et une « excroissance » émergent.

**Figure 7.3**

Le ribosome 70S peut se maintenir par associations entre des régions particulières des deux sous-unités, comme le montrent ces deux images.

INTERACTIONS ENTRE LES PROTÉINES RIBOSOMALES ET L'ARNr

Les ARN constituent la majeure partie de la masse du ribosome. Leur présence est « insinuante » et probablement presque toutes les protéines ribosomales sont associées à l'ARNr. Ainsi, les ARNr majeurs forment ce qui est parfois perçu comme le squelette de chaque sous-unité, une armature continue dont la présence domine la structure et qui détermine les positions des protéines ribosomales.

**Figure 7.4**

Les formes des ribosomes bactériens et des sous-unités sont révélées par la microscopie électronique.

Un exemple de chaque structure est montré parallèlement à son modèle. Photographies aimablement fournies par James Lake.

Figure 7.5

Les comparaisons de taille montrent que les ribosomes sont assez grands pour accueillir deux ARNt (ainsi que 40 bases de l'ARNm).

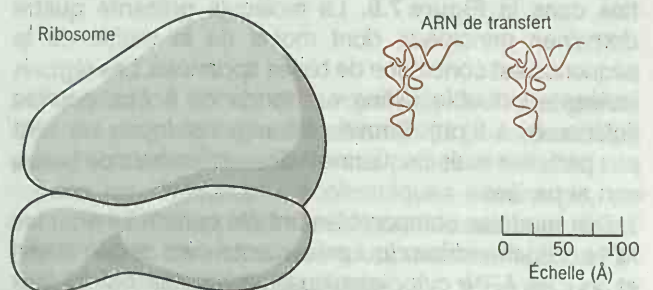


Tableau 7.4
Les longueurs des ARNr majeurs présentent de grandes variations.

Origine du ribosome	Grand ARNr	Petit ARNr
Mitochondrie de trypanosome	1 230 bp	640 bp
Mitochondrie de mammifère	1 559 bp	854 bp
Mitochondrie de levure	3 273 bp	
Bactéries	2 904 bp	1 542 bp
Cytoplasme de levure	3 750 bp	2 000 bp
Cytoplasme de la mouche du vinaigre	4 100 bp	2 000 bp
Cytoplasme de mammifère	4 712 bp	1 950 bp

Comme le résume le **Tableau 7.4**, la taille des ARN ribosomiaux est très variable. Inférieure à 2 000 bases dans la mitochondrie du trypanosome, elle peut atteindre plus de 7 000 bases dans le cytoplasme de mammifère.

Les deux ARNr majeurs ont un nombre très élevé de bases appariées, formant sans doute des structures secondaires spécifiques. Bien que la séquence d'un ARN puisse être utilisée pour prévoir la formation de régions à bases appariées, en ce qui concerne des molécules aussi grandes que les ARNr, plusieurs configurations possibles peuvent exister. Il est impossible de prévoir laquelle serait retenue même si la molécule est simplement libre en solution. En tenant compte des stabilités relatives des bases appariées particulières, différents modèles peuvent être distingués ; mais ce genre de données est d'un intérêt limité.

L'approche la plus productive dans l'analyse de la structure secondaire est la comparaison des séquences d'organismes apparentés. Ces régions, importantes dans la structure secondaire, sont conservées ; ainsi, si une paire de bases est importante, elle peut se former dans chaque ARNr, à la même position relative. A partir de telles comparaisons, des modèles ont été élaborés pour les ARNr 16S et 23S.

Le modèle habituel de l'ARNr 16S d'*E. coli* est schématisé dans la **Figure 7.6**. La molécule présente quatre domaines principaux dont moins de la moitié de la séquence est constituée de bases appariées. Les régions isolées en double hélice ont tendance à être courtes (inférieures à 8 pb). Souvent les régions duplex ne sont pas parfaites mais contiennent des renflements de bases non appariées.

Des modèles comparables ont été construits pour les ARNr mitochondriaux (qui présentent moins de domaines) et pour les ARNr cytoplasmiques eucaryotes (qui présen-

tent plus de domaines). La longueur accrue des ARNr eucaryotes est due en grande partie à l'adjonction de séquences qui représentent les domaines additifs.

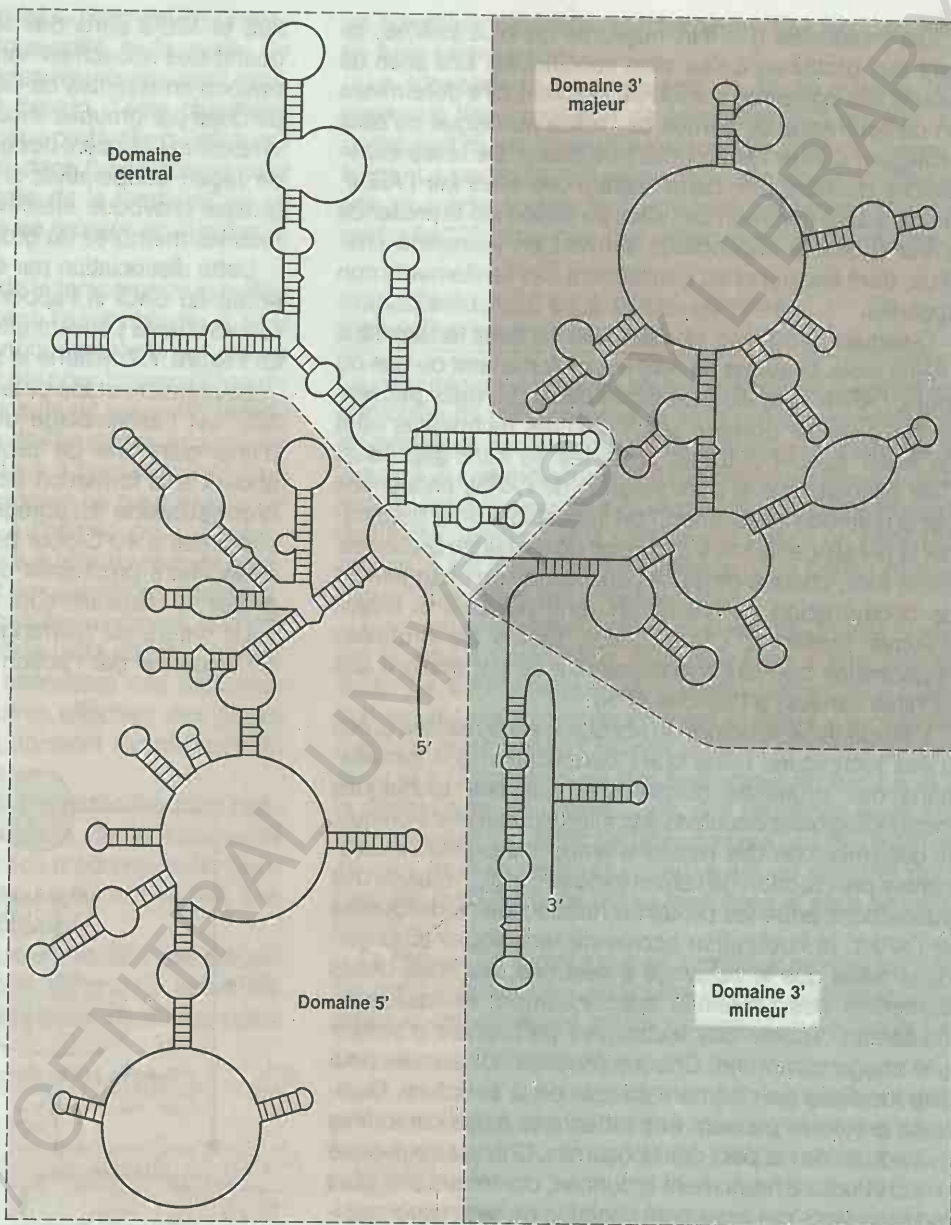
Ce modèle représente la structure d'un ARNr libre ; il ne tient pas compte des protéines ribosomiales dont la liaison avec l'ARN pourrait jouer sur la disponibilité des régions pour l'appariement de bases. Dans quelle mesure cette structure correspond-elle à l'organisation de l'ARNr dans la sous-unité ribosomale ?

Les méthodes qui permettent d'étudier la structure dans la sous-unité comprennent : la détermination des bases qui peuvent réagir avec le kethoxal (composé qui attaque les guanines non appariées), des régions qui s'entrecroisent après traitement au psoralène, des sites préférentiels de clivage par les nucléases et des régions de l'ARNr qui peuvent se lier à des protéines données. Jusqu'ici, les résultats sont relativement compatibles avec le modèle de structure de l'ARNr libre, mais il semble y avoir moins de bases appariées dans la sous-unité que dans la forme libre.

La structure de l'ARNr dans la sous-unité est-elle invariable ? Quelques différences dans la réactivité de l'ARNr 16S sont révélées quand les sous-unités 30S sont comparées aux ribosomes 70S ; il existe aussi quelques disparités entre les ribosomes libres et ceux qui sont engagés dans la synthèse protéique. Ceci suggère que des changements de conformation de l'ARNr peuvent apparaître quand l'ARNm est lié, quand les sous-unités s'associent ou quand l'ARNt est lié. Nous ne savons pas si de tels changements sont le reflet d'une interaction directe de l'ARNr avec l'ARNm ou l'ARNt, ou s'ils résultent indirectement d'autres changements dans la structure du ribosome. Il ressort principalement que la conformation du ribosome peut être flexible au cours de la synthèse protéique.

La structure primaire de l'ARNr est caractérisée par la présence de résidus méthylés, quelquefois dans des régions bien conservées. Il y a environ 10 groupes méthyle dans l'ARNr 16S (situés pour la plupart vers l'extrémité 3' de la molécule) et sans doute environ 20 dans l'ARNr 23S. Dans les cellules de mammifères, les ARNr 18 et 28S contiennent respectivement 43 et 73 groupes méthyle, ainsi environ 2 % des nucléotides sont méthylés (soit trois fois plus que chez les bactéries).

La grande sous-unité ribosomale contient aussi une molécule de 120 bases, l'ARNr 5S (présente dans tous les ribosomes à l'exception des ribosomes mitochondriaux). Dans les ARNr 5S procaryotes, quelques séquences sont particulièrement conservées dans les régions qui se lient aux protéines ribosomiales. De la même façon,

**Figure 7.6**

Les quatre domaines de l'ARNr 16S : chacun contient plusieurs courtes régions double-brin.

les ARN 5S eucaryotes sont conservés, avec, par exemple, une séquence qui prédomine chez les mammifères. Cependant, il n'y a pas de relation entre les séquences procaryotes et eucaryotes.

Toutes les molécules d'ARN 5S présentent une structure dont la plus grande partie est constituée de bases appariées, bien qu'il ait été difficile d'établir un modèle

exact (plus de 20 ont été proposés). Ceci souligne la difficulté pour distinguer les alternatives possibles d'appariement, même dans de toutes petites molécules.

Dans les ribosomes cytoplasmiques eucaryotes, un autre petit ARNt apparaît dans la grande sous-unité. C'est l'ARNr 5,8S. Sa séquence peut correspondre à l'extrémité 5' de l'ARNr 23S procaryote.

Les molécules d'ARNr, majeures ou plus petites, se lient aux protéines à des sites spécifiques. Les sites de liaison des protéines sur l'ARNr peuvent être déterminés en caractérisant les parties de l'acide nucléique qu'elles protègent contre l'action des nucléases. De telles expériences donnent une carte linéaire des sites sur l'ARNr. Un caractère commun des sites de liaison est la présence d'une structure secondaire, souvent en épingle à cheveux, dont les branches contiennent des renflements non appariés.

Quelques protéines ribosomales se lient fortement à l'ARNr isolé. D'autres se lient moins fortement ou pas du tout à l'ARNr libre isolé de façon classique mais peuvent se lier à l'ARNr préparé selon d'autres techniques ou à un ARNr auquel d'autres protéines se sont déjà liées. Ceci suggère que la conformation de l'ARNr (influencée par la méthode d'isolement) est importante pour déterminer la présence de sites de liaison de certaines protéines. En se liant, chaque protéine peut induire un changement de conformation de l'ARNr qui rend possible la liaison d'autres protéines. Chez *E. coli*, toutes les protéines ribosomales peuvent théoriquement se lier (malgré des affinités variées) à l'un des ARNr.

L'étude de la structure du ribosome a été réalisée grâce à des techniques telles que l'identification des localisations des protéines ribosomales par des techniques immunologiques couplées à la microscopie électronique, la détermination des relations entre des protéines adjacentes par réaction de liaison (crosslinking), l'analyse des interactions entre les protéines ribosomales individuelles et l'ARNr, et l'utilisation accrue de techniques biophysiques telles que le balayage à neutrons des sous-unités contenant des protéines spécifiquement marquées au deutérium. Toutes ces techniques permettent d'obtenir une image commune. Chaque protéine ribosomale peut être localisée précisément au sein de la structure. Quelques protéines peuvent être rattachées à des caractères individuels de l'aspect des ribosomes. Chaque sous-unité a une structure hautement ordonnée, contenant des sites indépendants qui assument chacune de ses responsabilités catalytiques.

LA RECONSTITUTION IN VITRO REPRODUIT L'ASSEMBLAGE IN VIVO

Quand des sous-unités ribosomales bactériennes sont centrifugées en CsCl, un groupe discret de protéines (les **protéines faiblement associées** ou « **split proteins** ») est perdu par chaque sous-unité. La dissociation engendre des particules « cœur » (cores) respectivement de

23S et 42S à partir des sous-unités 30S et 50S. En fait, quand des ribosomes sont mis en présence de concentrations croissantes de CsCl ou LiCl, des protéines sont perdues par groupes successifs. On peut en déduire que le ribosome contient des groupes de protéines organisées de façon coopérative, si bien que la dissociation d'un groupe provoque effectivement la perte simultanée de tous les membres du groupe.

Cette dissociation par étapes peut être inversée par le retrait du CsCl et l'apport d'ions Mg^{2+} . Les sous-unités reconstituées peuvent effectuer une synthèse protéique. La Figure 7.7 résume le procédé de reconstitution.

La première étape de la reconstitution d'une sous-unité 30S est l'assemblage entre l'ARNr 16S et un groupe d'une quinzaine de protéines. A froid, cette réaction aboutit à la formation de **particules intermédiaires de reconstitution** (en abrégé **IR**), particules qui doivent être chauffées à 40°C pour permettre un réarrangement unimoléculaire produisant des **particules IR***. Puis, les six protéines restantes sont reliées.

La nature du réarrangement unimoléculaire est mise en évidence par l'action de l'EDTA sur les sous-unités

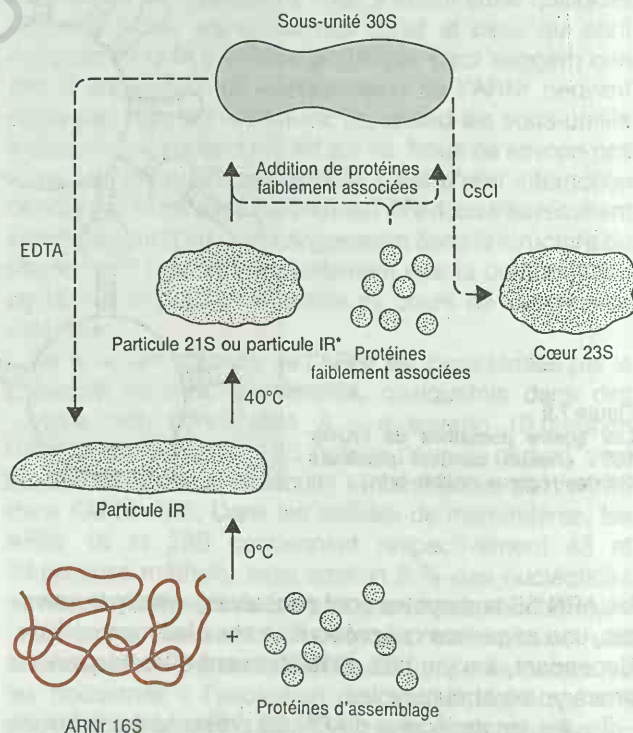


Figure 7.7

Les sous-unités 30S peuvent être dissociées en sous-particules qui sont proches des intermédiaires d'assemblage.

30S. L'agent chélateur en retenant le magnésium provoque un déploiement de la sous-unité et la perte de quelques protéines. Les particules résultantes sont similaires aux IR puisqu'elles ont besoin d'être chauffées avant de pouvoir se réassocier aux protéines qu'elles ont perdues. *Ainsi l'étape limitante dans l'assemblage peut être une conversion dépendante de la température au cours de laquelle les particules se replient en une structure plus compacte.*

Cette dépendance vis-à-vis de la température suggère que des mutations qui bloquent l'assemblage *in vivo* pourraient être détectées en tant que mutants **sensibles au froid**. De tels mutants défectifs pour l'assemblage de sous-unités (en anglais *sad* pour subunit assembly defective) peuvent être trouvés chez des bactéries incapables d'effectuer une synthèse protéique à basse température puisque l'assemblage des ribosomes ne peut plus avoir lieu. D'autres mutants, isolés à l'origine pour d'autres propriétés (telle que la résistance aux antibiotiques agissant sur le ribosome) présentent des altérations similaires.

Quand la température des cellules mutantes *sad* est abaissée, la production de la sous-unité affectée cesse. Certaines mutations sont des altérations des protéines ribosomales individuelles ; d'autres affectent des gènes dont les produits d'expression agissent directement ou indirectement sur l'assemblage.

Les mutants *sad* accumulent des particules plus petites, précurseurs dans l'assemblage de la sous-unité ribosomale affectée. Ces particules précurseurs contiennent l'ARNr associé à une partie seulement et non à la totalité des protéines de la sous-unité.

Quand l'assemblage des sous-unités 30S est bloqué un précurseur sédimente à 26S ou à 21S selon les conditions expérimentales. Il a été aussi identifié comme

un constituant mineur (et transitoire chez les bactéries de type sauvage).

Le blocage de l'assemblage de la sous-unité 50S provoque l'accumulation de particules 32S ou 43S. La particule 32S est la première étape identifiée de l'assemblage. La particule 43S représente une étape ultérieure ; elle possède toutes les protéines présentes dans la particule 32S ainsi que des espèces supplémentaires.

L'ASSEMBLAGE DES SOUS-UNITÉS EST LIÉ À LA TOPOLOGIE

Le précurseur de la petite sous-unité trouvé *in vivo* et les particules IR reconstituées *in vitro* ont une constitution en protéines proche (ou identique) comme l'indique la Figure 7.7. Les molécules d'ARN sont sensiblement différentes, car le précurseur *in vivo* a une forme précurseur de l'ARNr ; tandis que la particule *in vitro* a un ARN 16S mature. Il est intéressant (et heureux) que l'ARNr mature puisse être utilisé pour l'assemblage.

L'unique précurseur de la sous-unité 30S trouvé *in vivo* est le seul intermédiaire distinct qui peut être mis en évidence. L'ARN précurseur, un peu plus long que l'ARN mature, est seulement légèrement méthylé. Le changement de conformation du précurseur est probablement associé à la suppression des séquences en surplus et à la méthylation de l'ARNr ; puis, les protéines restantes sont associées. L'ARNr 16S libre ne peut pas être méthylé *in vitro*, mais la particule « cœur » est un substrat pour l'activité méthylase, ce qui va bien dans le sens de cette hypothèse.

La Figure 7.8 montre que les protéines présentes dans les précurseurs et les particules IR sont également très proches des protéines qui disparaissent en dernier après

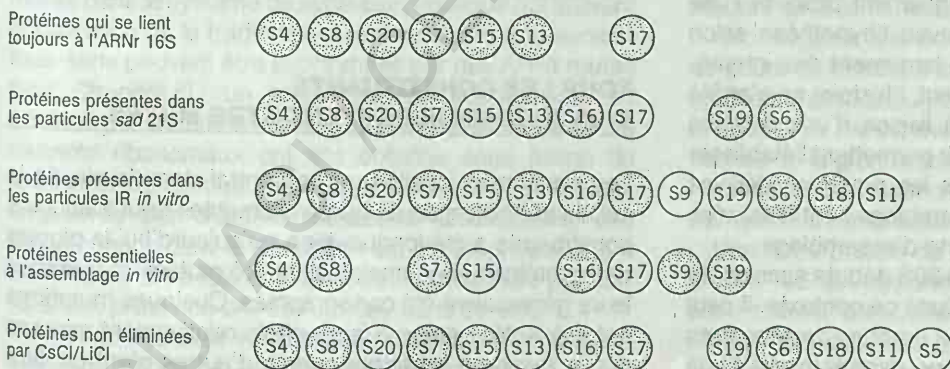


Figure 7.8
Les ensembles apparentés des protéines ribosomales sont présents dans les particules précurseurs assemblées *in vivo* ou *in vitro* et dans les particules cœur.

Les particules habituellement présentes sont indiquées par les cercles grisés : celles qui apparaissent moins souvent sont indiquées par les cercles blancs.

traitement au LiCl ou au CsCl. Ceci laisse supposer que la topologie de la sous-unité 30S est le reflet du mécanisme d'assemblage. Ces protéines qui font partie du précurseur tendent à être les composants les plus fixes de la sous-unité 30S dont l'élimination, plus difficile, nécessite un traitement énergétique.

Les rôles des composants particuliers dans l'assemblage peuvent être analysés en essayant de reconstituer les sous-unités en l'absence des protéines individuelles. Il n'y a pas de reconstitution quand certaines protéines sont absentes du mélange d'assemblage. Elles sont donc essentielles à l'assemblage. D'autres protéines peuvent être omises ; elles font partie des espèces ajoutées en dernier et dont la présence n'est pas nécessaire pour déterminer la structure globale. Les sous-unités qui s'assemblent peuvent avoir des caractéristiques biophysiques normales, alors que testées en vue d'une synthèse protéique, elles sont inopérantes dans certaines fonctions (révélant ainsi le rôle de la protéine absente).

La reconstitution partielle peut être utilisée pour découvrir si l'absence d'une protéine affecte la capacité des autres protéines à s'assembler dans la particule. La technique consiste essentiellement à éliminer une protéine et à regarder quelles sont les autres protéines qui ne peuvent plus se lier. (Ceci peut être réalisé sans que l'assemblage partiel produise nécessairement une sous-unité fonctionnelle).

La carte d'assemblage *in vitro* de la Figure 7.9 a été établie en déterminant quelles protéines doivent être liées à l'ARNr 16S pour permettre l'association d'une autre protéine. Le résultat met en évidence que les protéines ribosomales doivent se lier à la particule en cours d'assemblage selon un certain ordre. Certaines des protéines qui montrent des relations dépendantes dans la carte d'assemblage sont en fait physiquement proches l'une de l'autre. Ceci est conforme avec l'hypothèse selon laquelle la dépendance implique largement des interactions physiques ; mais, évidemment, l'histoire ne s'arrête pas là, puisque (par exemple) la liaison d'une protéine peut ouvrir un autre site sur l'ARNr permettant l'établissement d'une autre liaison. Toutes les protéines définies comme étant essentielles à l'assemblage sont localisées dans la première partie de la carte d'assemblage.

L'assemblage de la sous-unité 30S débute sans doute tandis que l'ARNr est transcrit. Dans ce contexte, il peut être significatif que le plus grand nombre de sites forts de fixation de l'ARNr soit regroupé dans la partie 5' de la molécule qui est transcrite la première. Il est probable que l'ARNr qui se détache du précurseur soit déjà relié à plusieurs protéines ribosomales comme l'illustre la Figure 7.10.

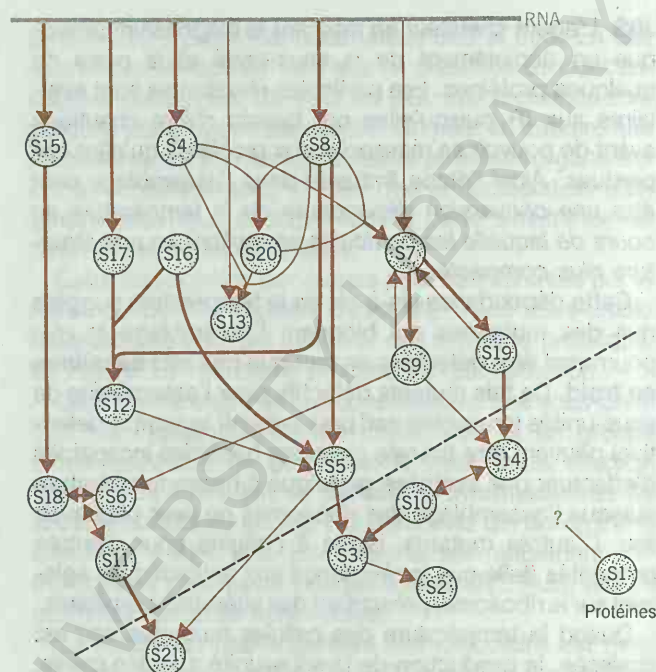


Figure 7.9

L'assemblage du ribosome fait intervenir une série de réactions séquentielles.

La flèche qui relie deux protéines indique que la première aide à la liaison de la seconde (une double flèche montre les effets mutuels). Les effets majeurs sont indiqués par les lignes épaisses alors que les lignes fines indiquent les effets mineurs. Les flèches issues de l'ARNr montrent les protéines liées le plus fortement, mais d'autres sont également liées.

Les protéines les plus proches de l'ARNr sont celles qui persistent dans les particules précurseurs ; celles qui en sont plus éloignées s'assemblent plus tard dans la particule et sont éliminées plus facilement par les sels. La ligne en pointillés sépare les protéines qui sont habituellement ou souvent trouvées dans les particules précurseurs de celles qui ne sont jamais trouvées (équivalentes aux protéines faiblement associées).

TOUS LES CONSTITUANTS D'UN RIBOSOME PEUVENT ÊTRE MUTÉS

Les ribosomes jouent un rôle central dans la vie de la cellule et l'obtention de bactéries mutées pour tous leurs constituants a été longue, bien qu'aujourd'hui la plupart des protéines ribosomales aient été mutées et que tous leurs gènes aient été cartographiés. Quelques mutations ont été isolées grâce à leurs effets relativement mineurs sur la synthèse protéique ; mais d'autres n'ont pu être isolées qu'en ayant recours à l'utilisation de létalité conditionnelle.

Certains antibiotiques agissent en inhibant la synthèse protéique bactérienne (voir Tableau 5.6). L'obtention de

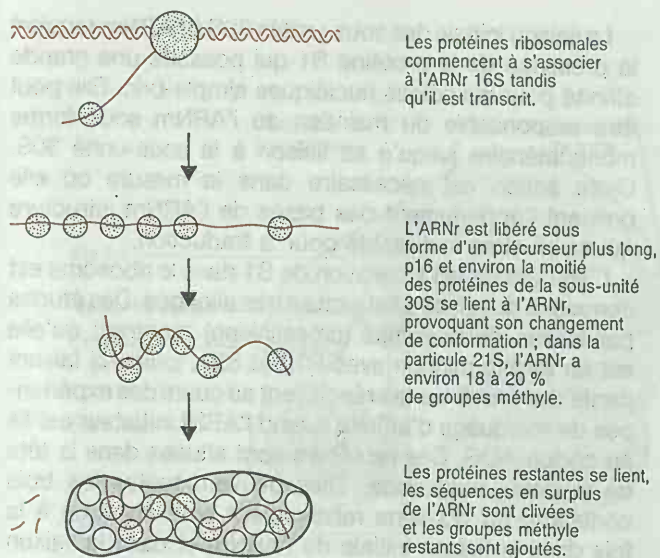


Figure 7.10

L'assemblage de la sous-unité du ribosome est un processus ordonné au cours duquel des changements de conformation apparaissent à chaque addition de protéine.

mutants qui résistent à leur action est possible. Les mutations modifient souvent les protéines ribosomales (ou les enzymes qui agissent sur les constituants du ribosome). Cinq protéines 30S et quatre protéines 50S ont été mutées pour conférer une résistance aux antibiotiques. Ces variants sont utiles pour définir le rôle de la protéine dans le ribosome.

Des essais directs d'isolement de mutants pour la synthèse protéique sont possibles à partir de changements dans le système de synthèse protéique qui altèrent la précision de la traduction. Des mutations non-sens et faux-sens peuvent être supprimées par des ARNt mutés (voir Chapitre 6) ; un autre type de suppression est conféré par mutation dans le ribosome. Ainsi plusieurs mutants ribosomiaux ont été obtenus sous forme de revertants du second type qui passent outre les déficits engendrés par d'autres mutations. Quelquefois, une étape connue dans la traduction est affectée, permettant de déterminer le rôle d'une protéine ribosomale particulière. Six protéines ont été mutées dans ce but.

Nous avons déjà vu que des mutants dans l'assemblage du ribosome peuvent être obtenus grâce à leur sensibilité à la température. Ce phénotype appartient aussi à certains mutants résistants aux antibiotiques. Jusqu'ici, quatre des protéines du 30S et une protéine du 50S ont été identifiées dans des mutants d'assembla-

ge. Ce type de mutation est conditionnellement létal puisque, sans ribosomes, la cellule meurt.

Quelques protéines ribosomales ont été identifiées simplement en raison de leur altération chez des mutants sensibles à la température. Ceci signifie que la diminution de croissance de bactéries à haute température résulte de l'altération d'une protéine ribosomale qui (probablement) fonctionne mal quand la température s'élève. La généralité du déficit n'apporte pas d'indices sur la fonction de la protéine.

Il y a encore quelques protéines ribosomales pour lesquelles des mutations délétaires n'ont pas encore été trouvées. Il reste possible que ces protéines ne soient pas indispensables à l'assemblage et à l'activité du ribosome.

Les propriétés des mutants résistants à la kasugamycine établissent que l'ARNr 16S joue un rôle direct dans la synthèse protéique. La kasugamycine bloque la réaction d'initiation. Les mutants résistants de type *ksgA* sont dépourvus de méthylase qui apporte quatre groupements méthyle à deux adénines adjacentes en un site proche de l'extrémité 3' de l'ARNr 16S. Cette méthylation engendre une séquence hautement conservée $G - m_2^6A - m_2^6A$, retrouvée dans le petit ARNr procaryote et eucaryote.

Cependant, dans le type sauvage, la présence des quatre groupements méthyle rend le ribosome sensible à la kasugamycine. Pourquoi l'absence des groupements méthyle confère-t-elle une résistance à cet antibiotique ? Les sous-unités 30S sauvages et mutantes sont sensibles, de façon identique, à l'inhibition de liaison de la fMet-ARNt_i par la kasugamycine. Mais en présence des sous-unités 50S, les ribosomes sensibles (sauvages) libèrent le fMet-ARNt_i, contrairement aux mutants, permettant ainsi la continuité de la synthèse protéique.

Ceci suggère que les groupements méthyle sont actifs au cours de la liaison entre les sous-unités 30S et 50S ; réaction d'association qui doit aussi être connectée à la rétention de l'ARNt initiateur. Dans cette réaction, l'influence de la grande sous-unité est mise en évidence par l'isolement de mutations des protéines de la particule 50S qui rendent le ribosome sensible à la kasugamycine.

Une autre preuve de l'importance de l'extrémité 3' de l'ARNr 16S est apportée par sa sensibilité à la colicine E3. Produite par certaines bactéries, la colicine clive l'ARNr 16S d'*E. coli* à environ 50 nucléotides de son extrémité 3'. Ce clivage inhibe totalement l'initiation de la synthèse protéique.

Le chloramphénicol bloque la synthèse protéique en interférant sur la peptidyl transférase. La résistance apparaît de façon tout à fait différente chez les bactéries et

les mitochondries. Chez les bactéries, les protéines du 50S des mutants sont altérées. Dans les mitochondries, il s'agit de changements simples de bases à environ 243 nucléotides de l'extrémité 3' du grand ARNr.

LES RIBOSOMES ONT PLUSIEURS CENTRES ACTIFS

Plusieurs activités ribosomales, sur la base de leur rôle ou de leur localisation, peuvent être distinguées. Les sous-unités 30S se lient à l'ARNm et au complexe ARNt initiateur-facteur d'initiation ; elles s'unissent ensuite aux sous-unités 50S. Les ribosomes 70S possèdent les sites A et P, fonctionnels, auxquels l'ARNt est lié ; le centre peptidyl-transférase est porté par la sous-unité 50S. Le site de liaison EF-G, et par là-même la possibilité de translocation, est également porté par la sous-unité 50S.

Une représentation simplifiée de ces sites est illustrée par la **Figure 7.11**. Ils constituent les deux tiers environ de la structure ribosomale, reconnus comme le **domaine traductionnel**. Le tiers restant du ribosome représente le **domaine de sortie** par l'intermédiaire duquel les ribosomes sont liés aux membranes (voir Chapitre 8). Une chaîne polypeptidique passe d'un domaine à l'autre à travers un site de sortie situé à distance de la peptidyl-transférase.

Comment ces sites sont-ils reliés à la véritable topologie du ribosome ? Quels sont les constituants du ribosome qui sont impliqués dans ses diverses fonctions ? L'analyse de la relation structure-fonction a nécessité l'emploi de plusieurs techniques.

Un marquage d'affinité est un moyen d'identification de sites particuliers. Cette technique utilise les analogues des constituants qui se lient au ribosome. Ils sont soit eux-mêmes potentiellement actifs, soit activés *in situ*. Par exemple, l'utilisation d'ARNt comportant des modifications appropriées permet d'identifier les éléments du ribosome auxquels un marquage a été transféré.

La modification de sous-unités entières a été largement utilisée. Un moyen d'approche consiste à traiter les sous-unités avec des réactifs qui dénaturent les protéines ou l'ARN et à établir ensuite la relation entre une dégradation particulière et la perte spécifique d'une fonction. Par exemple, le kethoxal a été utilisé pour modifier les guanines de l'ARNr et le tétranitrométhane pour nitrater les protéines. Afin que la perte d'une fonction puisse être étroitement corrélée à une fonction particulière, la limite de la dégradation doit être bien maîtrisée.

La liaison initiale des sous-unités 30S à l'ARNm requiert la présence de la protéine S1 qui possède une grande affinité pour les acides nucléiques simple-brin. Elle peut être responsable du maintien de l'ARNm sous forme monocaténaire jusqu'à sa liaison à la sous-unité 30S. Cette action est nécessaire dans la mesure où elle prévient l'appariement des bases de l'ARNm, structure qui serait alors inutilisable pour la traduction.

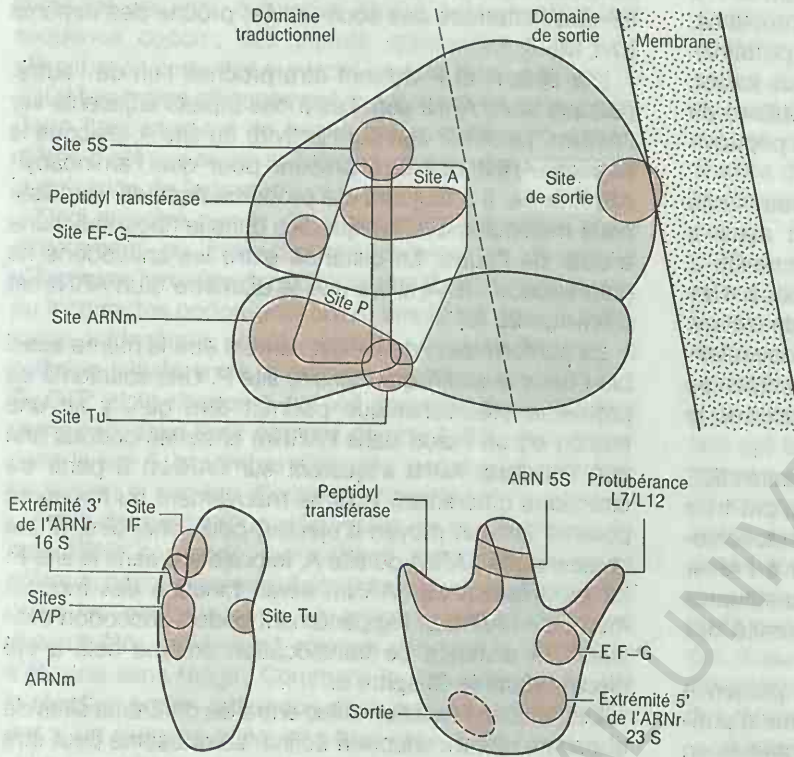
L'analyse de la localisation de S1 dans le ribosome est complexe du fait de sa structure très allongée. Des études par liaison de proximité (crosslinking) montrent qu'elle est en étroite relation avec S18 et S21, celles-ci faisant partie des protéines qui réagissent au cours des expériences de marquage d'affinité quand l'ARNt initiateur est lié au codon AUG. Ces protéines sont situées dans la tête de la petite sous-unité. Elles peuvent toutes les trois constituer un domaine relativement petit impliqué à la fois dans la liaison initiale de l'ARNm et dans la liaison de l'ARNt initiateur.

L'extrémité 3' de l'ARNr 16S, située également non loin de ce domaine, joue un rôle dans la réaction de liaison. Elle possède une courte séquence complémentaire d'une séquence présente avant chaque codon d'initiation AUG dans l'ARNm bactérien. Il est probable que l'appariement de bases complémentaires entre ces séquences intervienne au cours de la liaison initiale des ribosomes aux sites d'initiation sur l'ARNm (voir Chapitre 8).

Les facteurs d'initiation se lient dans la même région du ribosome. IF3 peut être lié de façon croisée à l'extrémité 3' de l'ARNr, aussi bien qu'à plusieurs protéines ribosomales, y compris celles qui sont probablement impliquées dans la liaison de l'ARNm. Si cette partie de la petite sous-unité est aussi celle qui est engagée dans la liaison à la grande sous-unité, le rôle d'IF3 pourrait être de se lier à ce niveau pour stabiliser l'ensemble, ARNm-sous-unité 30S, puis d'être déplacé lors de la jonction avec la sous-unité 50S.

Le mécanisme de jonction des sous-unités n'est pas encore totalement élucidé. Il existe une certaine complémentarité entre une partie de l'ARNr 23S et (à nouveau) la région 3' de l'ARNr 16S. Une éventualité est que l'appariement de bases serve à rassembler les sous-unités. Dans ce cas, la capacité d'IF3 à se lier à l'ARNr 16S pourrait être incompatible avec l'association des sous-unités.

L'ARNr 16S peut faire partie du site A. Des études par liaison de proximité ont mis en évidence que le peptidyl-ARNt peut être uni en un point proche de l'extrémité 3' de l'ARNr 16S. La séquence conservée de l'ARNr peut interagir avec l'ARNt. Cependant, aucune liaison de

**Figure 7.11**

Le ribosome possède plusieurs centres actifs séparés.

Les sous-unités qui sont montrées ont subi une rotation de 90° par rapport au ribosome intact. Les localisations des divers sites sont strictement schématisiques et ne peuvent pas être reliées à la structure précise en trois dimensions du ribosome.

Le *site de liaison de l'ARNm* est situé dans la sous-unité 30S et comprend les protéines S1, S18 et S21, aussi bien que S3, S4, S5, S12 et la région terminale en 3' de l'ARNr 16S ; c'est aussi le site de liaison des facteurs IF. Il est proche du site P.

Le *site P* est probablement localisé en grande partie dans la sous-unité 30S et est capable de lier l'initiateur ARNt. L'ARNr 16S est un constituant faisant sans doute intervenir une région proche de son extrémité 3'. Sur la sous-unité 50S, le site P inclut L2 et L7, de même que L14, L18, L24 et L33.

Le *site A* doit être proche du site P mais la géométrie exacte est inconnue. L'ARNr 16S a une influence sur le site A et pourrait en être un constituant. La plupart des protéines identifiées avec le site A sont sur la sous-unité 50S ; elles comprennent L1, L5, L7/L12, L20, L30 et L33.

Le *site peptidyltransférase* devrait se situer quelque part dans la région reliant les sites A et P, proche de la fin de l'ARNt. Les protéines nécessaires à cette activité sont L2, L3, L4, L15 et L16.

Le *site ARN 5S* se trouve sur la protubérance centrale de la grande sous-unité et comprend L5, L18 et L25.

Le *site Tu* est à la surface externe de la petite sous-unité et comprend plusieurs protéines proches les unes des autres, qui peuvent être liées de façon croisée à EF-Tu.

Le *site L7/L12* forme la protubérance de la sous-unité 50S qui est constituée de deux dimères.

Le *site EF-G* peut être localisé sur la grande sous-unité près de l'interface avec la petite sous-unité, près de S12.

proximité ne peut être obtenue entre l'ARNr et l'aminocyl-ARNt au site A.

Des études par liaison de proximité concernant EF-Tu montrent que le site A est voisin d'un groupe de protéines, mais nous ne savons pas encore comment ces protéines (et d'autres) interagissent pour former le site de liaison de l'ARNt. De la même façon, un groupe de protéines du site P a été identifié autour de l'extrémité 3' du peptidyl-ARNt.

Le site de liaison de EF-Tu se trouve sur la sous-unité 30S, du côté opposé à la zone de contact avec la sous-unité 50S. Il est contigu au site de liaison de EF-G localisé près de l'interface entre les sous-unités. Ces positions relatives prouvent que les sites de liaison peuvent être étendus, chacun occupant une place relativement importante de la structure du ribosome. Elles ne sont pas de petites régions indépendantes comme le sont les centres actifs enzymatiques.

L'incorporation de l'ARN 5S dans les sous-unités 50S assemblées *in vitro* dépend de la propriété qu'ont trois protéines L5, L8 et L25 de former un complexe stœchiométrique avec lui. Ce complexe peut se fixer à l'ARNr 23S, mais, individuellement, aucun de ses constituants ne peut le faire. Le complexe se situe à proximité des sites A et P.

L'activité peptidyl transférase nécessite la présence d'un groupe de protéines. Il n'a pas été possible d'attribuer l'activité enzymatique à l'une de ces protéines en particulier. Cependant, le site peptidyl transférase a été localisé, grâce à la liaison de la puromycine, sur la protubérance centrale. Il semble aussi qu'outre les protéines, l'ARNr 23S soit impliqué dans la création du site peptidyl transférase.

La chaîne polypeptidique naissante paraît s'extraire de la sous-unité 50S à environ 150 Å du site peptidyl transférase. Elle s'étend probablement à travers le ribosome sous la forme d'une chaîne polypeptidique non repliée jusqu'à ce qu'elle quitte le domaine de sortie, libre ensuite d'amorcer son repliement.

L'unique exception à l'unimolarité des protéines ribosomales est illustrée par L7/L12. L7 ne diffère de L12 que par la présence d'un groupement acétyl sur le N-terminal ; il existe deux copies de ce dimère par ribosome. L'agrégat L7/L12 forme la tige de la grande sous-unité. Quand il disparaît, les particules sont incapables d'hydrolyser le GTP. Cela ne signifie pas nécessairement que L7/L12 est la GTPase ; il peut en revanche être nécessaire à l'activité d'une autre protéine.

Le facteur de translocation EF-G se lie à la sous-unité 50S ; mais la translocation sous-entend un mouvement

de l'ARNm à travers la sous-unité 30S. Le site de liaison d'EF-G est proche de S12, l'une des protéines du site de liaison de l'ARNm sur la sous-unité 30S. Ceci situe EF-G à l'interface des sous-unités, proche des dimères L7/L12.

Les sites A et P doivent être proches l'un de l'autre, puisque leurs ARNt sont liés à des triplets adjacents sur l'ARNm. Le site P agit sur l'activité du site A, puisque le peptidyl-ARNt doit être présent pour que l'aminocyl-ARNt se lie. Il a toujours été problématique de visualiser deux molécules d'ARNt ajustées dans le ribosome, l'une à côté de l'autre. La distance entre les anticodons ne peut excéder 10 Å, alors que le diamètre d'un ARNt est d'environ 20 Å.

La conformation de l'ARNt semble être la même aussi bien dans le site A que dans le site P. Une solution à ce problème stéréochimique pourrait être qu'il y ait une torsion ou un nœud dans l'ARNm entre les codons afin que les deux ARNt s'ajustent sur l'ARNm à partir de directions différentes. Puis, le mouvement du ribosome pourrait être un moyen d'éjecter, pour ainsi dire, d'une chiquenaude l'ARNt du site A, le projetant dans le site P. La progression de l'ARNm serait fonction des mouvements de l'ARNt et l'appariement codon-anticodon évaluerait la distance de translocation comme cela a été discuté dans le Chapitre 6.

Les relations fonctionnelles entre les différents sites du ribosome restent encore à définir. Le ribosome peut être une structure puissamment interactive, à l'intérieur de laquelle un changement ponctuel pourrait grandement affecter l'activité d'un autre site.

LA PRÉCISION DE LA TRADUCTION

L'absence de variation décelable au cours de l'analyse de séquence d'une protéine démontre que la synthèse protéique doit être extrêmement précise : très peu d'erreurs apparaissent sous forme de substitution d'un acide aminé par un autre. Il y a deux étapes de la synthèse protéique au cours desquelles des erreurs peuvent se produire.

Qu'un ARNt se charge du seul acide aminé approprié est évidemment critique. Nous avons vu dans le Chapitre 6 que l'aminocyl-ARNt-synthétase remplit cette fonction. Le risque d'erreur varie sans doute en fonction de chaque enzyme, mais on estime habituellement que les erreurs se produisent dans moins d'une aminocyclation sur 10⁵.

Une des caractéristiques troublantes de la synthèse protéique est sa fidélité. Bien que les constantes de

liaison varient avec la réaction individuelle codon-anticodon, la spécificité est toujours beaucoup trop faible pour fournir un taux d'erreur de 10^{-5} . En solution, les ARNt libres ne se lient que relativement faiblement à leur séquence codon ; des triplets apparentés mais faux (deux bases correctes sur trois) ne sont reconnus que 10 à 100 fois moins efficacement que les codons appropriés. Donc l'appariement de bases codon-anticodon en lui-même n'est pas assez spécifique pour rendre compte de la précision de la traduction.

Ceci suggère que le ribosome agit en quelque sorte directement ou indirectement comme un correcteur d'épreuves (proofreader), distinguant les paires correctes ou incorrectes codon-anticodon, amplifiant ainsi la différence intrinsèque plutôt modeste. Supposons que la collision initiale entre le complexe aminoacyl-ARNt.EF-Tu.GTP et le ribosome ne soit pas spécifique. Si tout complexe, sans tenir compte de son ARNt, peut entrer dans le site A, le nombre d'entrées incorrectes excèdera largement le nombre d'entrées correctes.

Il doit donc exister un mécanisme qui stabilise l'aminocyl-ARNt approprié, permettant à son acide aminé d'être accepté comme substrat de la chaîne polypeptidique ; les contacts avec un aminoacyl-ARNt incorrect doivent être rapidement rompus afin que le complexe s'éloigne sans réagir. Comment le ribosome évalue-t-il la réaction codon-anticodon dans le site A pour déterminer si un ajustage adéquat a bien été réalisé ?

La capacité du ribosome à influencer la précision de la traduction a d'abord été mise en évidence grâce aux effets des mutations qui confèrent la résistance à la streptomycine. Un des effets de la streptomycine est l'augmentation du nombre de fausses lectures des pyrimidines U et C (habituellement l'une est utilisée à la place de l'autre, éventuellement à la place de A). La streptomycine pourrait agir au niveau de la protéine S12 ; la séquence de cette protéine est altérée chez les mutants résistants. Les ribosomes à protéine S12 dérivée de bactéries résistantes ont un taux de fausses lectures diminué par rapport aux ribosomes de type sauvage. Ceci compense l'action de la streptomycine sur la fausse lecture.

Des mutations en deux autres loci qui codent pour les protéines S4 et S5 influencent la fausse lecture puisque des revertants qui présentent le niveau habituel de fausse lecture peuvent être isolés à l'un et l'autre locus. Ainsi la précision de la traduction est contrôlée par les interactions de ces trois protéines. Le niveau de fausse lecture et la nature de la réponse à la streptomycine dépendent tous les deux des versions de ces protéines présentes. (Dans certaines combinaisons, le ribosome est même

dépendant de la présence de la streptomycine pour traduire correctement.)

Un mécanisme direct de contrôle de la précision serait d'établir la stéréochimie du site A qui pourrait varier afin de déterminer la latitude de reconnaissance codon-anticodon. A l'origine les résultats obtenus avec la streptomycine ont été interprétés en terme de géométrie du ribosome, en supposant que la liaison codon-anticodon est étudiée d'une façon telle que le critère d'acceptation de l'aminocyl-ARNt pourrait être rendu plus ou moins précis.

Une alternative est que l'effet soit indirect. La vitesse de réaction du ribosome déterminerait le temps disponible à la reconnaissance de l'ARNt, et donc l'efficacité du mécanisme. Ce modèle pourrait expliquer l'effet de la streptomycine qui interviendrait sur la cinétique de l'élongation des chaînes polypeptidiques. Le paramètre important est la vitesse d'action du ribosome par rapport au temps requis pour créer et briser les contacts. Si la vitesse de formation de la liaison peptidique augmente, les aminoacyl-ARNt incorrects tendront plus à être captés dans la formation de la liaison avant que l'aminocyl-ARNt ne s'échappe. Le ralentissement du taux de synthèse protéique autorise un temps plus long dans la correction des erreurs. Il existe une certaine relation entre la vitesse d'élongation de la chaîne polypeptidique et la fréquence des fausses lectures.

L'hypothèse d'un contact correct entre le codon et l'anticodon qui pourrait être perçu à l'autre bout de l'ARNt par un changement de conformation peut être émise. Ce changement pourrait être nécessaire à l'acide aminé qui, en se localisant de façon appropriée, accepterait le polypeptide du peptidyl-ARNt.

Un problème important dans le calcul du coût de la synthèse protéique est de connaître l'étape où la décision d'accepter l'ARNt est prise. Si une décision immédiate est prise de libérer le complexe aminoacyl-ARNt.EF-Tu.GTP, il en coûte peu de rejeter le grand nombre d'ARNt incorrects susceptibles (statistiquement) d'entrer dans le site A avant que l'ARNt approprié ne soit reconnu. Mais si le GTP est hydrolysé quand le complexe se lie, une liaison riche en énergie supplémentaire sera clivée pour toute association incorrecte avec un ARNt. Ceci augmenterait le coût de la synthèse protéique bien au-delà des trois liaisons riches en énergie qui sont utilisées dans l'adjonction de chaque acide aminé (correct) à la chaîne. Il semble probable que l'utilisation du GTP *in vivo* est plus importante que prévue, comportant sans doute un clivage de 3 à 4 GTP supplémentaires par acide aminé.

POUR EN SAVOIR PLUS

Un ouvrage majeur de référence sur les ribosomes est le livre édité par **Nomura, Tissieres & Lengyel** : *Ribosomes* (Cold Spring Harbor Lab., New York, 1974). Bien qu'il soit dépassé, il fournit une vue générale très utile sur les aspects structuraux et fonctionnels. La structure du ribosome a été reprise de façon homogène par **Lake** (*Ann. Rev. Biochem.* **54**, 507-530, 1985).

L'ARN ribosomal a été étudié en détail par **Woese et al.** (*Microbiol. Rev.* **47**, 621-699, 1983) et **Noller** (*Ann. Rev. Biochem.* **53**, 119-162, 1984). Les structures et fonctions des sites ribosomiaux ont été décrits par **Pongs** (pp. 78-104 dans Altman (Ed), *Transfer RNA*, MIT Press, Cambridge, 1978). Les paramètres qui interviennent dans la précision de la traduction ont été brièvement résumés par **Kurland** (*Cell* **28**, 201-202, 1982).

CHAPITRE 8

L'ARN MESSAGEUR MATRICE

L'existence de l'ARNm fut initialement suspectée en raison de la nécessité d'un intermédiaire, dans les cellules eucaryotes, qui transporte l'information génétique du noyau où se situe l'ADN au cytoplasme où les protéines sont synthétisées. Conçu comme une matrice sur laquelle les acides aminés s'assembleraient en polypeptides, le messenger fut d'abord recherché chez les bactéries. Les ribosomes, reconnus comme site de synthèse protéique furent rapidement éliminés comme matrice possible, puis l'ARNm fut découvert en tant qu'espèce transitoire qui s'associe aux ribosomes pour être traduit en protéines (voir Chapitre 5).

Alors que l'existence transitoire de l'ARNm bactérien ne permettait pas l'isolement de la molécule, les propriétés du messenger étaient déduites, de façon détaillée, des caractéristiques de sa traduction en protéine. Ceci précédait de peu l'isolement de l'ARNm d'eucaryotes qui se révéla être un composant du cytoplasme relativement plus stable. Du fait de cette plus grande stabilité, l'ARNm eucaryote peut être isolé plus facilement ; aujourd'hui, on peut, en principe, isoler l'ARNm de chaque protéine.

Le rôle de l'ARNm est le même dans toutes les cellules vivantes : être traduit via le code génétique en protéines. Il existe cependant des différences dans les détails de la

synthèse et de la structure de l'ARN procaryote et eucaryote.

La disparité la plus évidente réside dans le fait que chez les eucaryotes l'ARNm est synthétisé sous forme d'une grande molécule précurseur dans le noyau. Après une étape de maturation complexe comportant souvent une réduction considérable de taille associée à d'autres modifications, l'ARNm est exporté vers le cytoplasme. Ainsi sa synthèse et son expression s'effectuent dans différents compartiments de la cellule.

Chez les bactéries, par contre, l'ARNm est transcrit et traduit dans l'unique compartiment cellulaire ; et les deux processus sont si intimement liés qu'ils s'effectuent simultanément. La différence principale dans la traduction est qu'un ARNm bactérien peut coder pour plusieurs protéines, alors qu'un ARNm eucaryote, invariablement, n'est traduit qu'en une seule chaîne polypeptidique.

LES MESSAGEURS BACTÉRIENS SONT TRANSITOIRES

Transcription et traduction sont étroitement liées chez les bactéries. Dès le début de la transcription, les ribosomes

s'attachent à l'extrémité 5' de l'ARNm et initient la traduction, avant même que le reste du message soit synthétisé. En effet un chapelet de ribosomes suit, immédiatement derrière, l'ARN polymérase.

La microscopie électronique de cellules d'*E. coli* a permis d'identifier les unités (inconnues) de transcription montrées dans la **Figure 8.1**, dans lesquelles plusieurs ARNm sont synthétisés simultanément, chacun portant les nombreux ribosomes engagés dans la traduction.

Transcription et traduction sont réalisées à des vitesses identiques. A 37 °C, l'ARNm est transcrit à raison de 2 500 nucléotides/minute, ce qui correspond à une synthèse d'environ 14 codons/seconde. Ceci est très proche de la vitesse de synthèse protéique, environ 15 acides aminés/seconde. Quand l'expression d'un nouveau gène est initiée, son ARNm apparaîtra dans la cellule en moins de 2 minutes et demie. La protéine correspondante apparaîtra, elle, peut-être dans les 30 secondes suivantes.

L'instabilité de la plupart des ARNm bactériens est frappante. La dégradation de l'ARNm suit de près sa traduction. Elle commence probablement dans la minute qui suit le début de la transcription. A nouveau, l'extrémité 5' de l'ARNm peut subir un début de dégradation avant que l'extrémité 3' ait été synthétisée ou traduite. La dégradation semble suivre le dernier ribosome du convoi le long de l'ARNm. Mais il est probable que la vitesse de dégradation soit deux fois moins élevée que celles de la transcription et de la traduction.

Figure 8.1.

Les unités de transcription peuvent être visualisées dans la bactérie.

La fine ligne centrale est l'ADN et les lignes qui en émergent sont les molécules d'ARN en cours de transcription. La longueur de la quinzaine d'ARN en voie de synthèse à partir de l'unité de transcription augmente de la gauche vers la droite, définissant la direction de la synthèse. Chaque ARN est recouvert de ribosomes.

Photographie aimablement fournie par Oscar Miller.

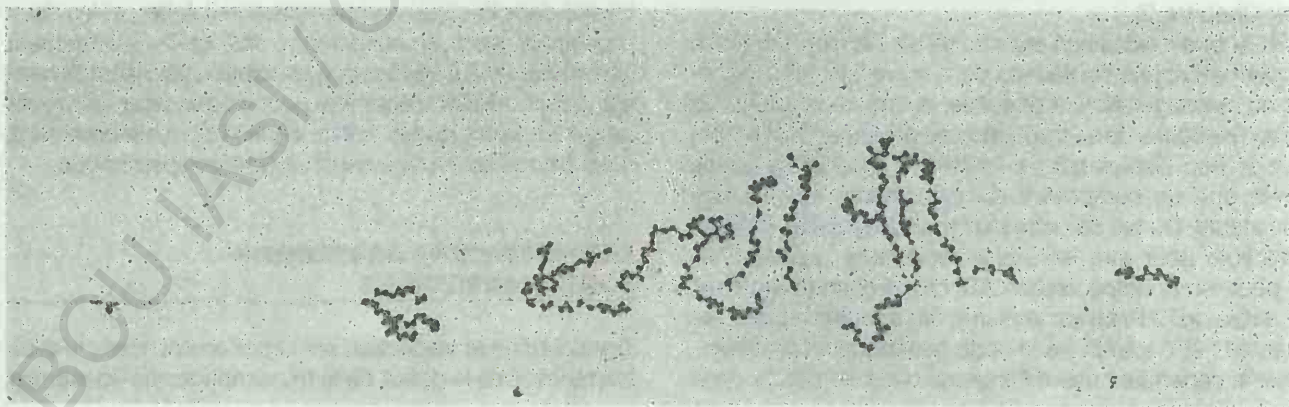
On peut mesurer la stabilité de l'ARN de deux façons. Elles reposent toutes les deux sur l'arrêt de la transcription de l'ARNm et le suivi du devenir des molécules d'ARNm dans la cellule.

La **demi-vie fonctionnelle** mesure la capacité de l'ARNm à servir de matrice pour la synthèse de sa protéine. Elle est habituellement d'environ 2 minutes. En d'autres termes, la quantité de nouvelle protéine qu'un ARNm particulier peut synthétiser diminue de moitié toutes les deux minutes.

La **demi-vie chimique** est déterminée en mesurant la diminution de la quantité d'ARNm capable de s'hybrider avec l'ADN de son gène. Généralement la décroissance chimique est légèrement postérieure au déclin fonctionnel.

Cette (discrète) différence suggère que la dégradation de l'ARNm comprend une étape initiale suffisante pour empêcher son utilisation en tant que matrice, mais indécélable sous forme d'une altération structurale qui entrave son hybridation. Par exemple, un seul clivage de l'ARNm pourrait ne pas influencer son hybridation avec l'ADN mais bloquer sa traduction.

L'étape initiale est suivie de la dégradation de l'ARNm en nucléotides qui le constituent. Les détails des mécanismes moléculaires responsables de la dégradation sont encore flous. Tout ce que nous savons est que la dégradation s'effectue plus ou moins séquentiellement le long de l'ARNm de l'extrémité 5' vers l'extrémité 3'.



Naturellement cet enchaînement de réactions n'est possible que parce que transcription, traduction et dégradation s'effectuent dans la même direction. La **Figure 8.2** donne une idée de la chronologie des événements pour l'expression d'une unité bactérienne de transcription typique.

N'essayons pas de visualiser le cas « moyen » qui rendrait encore plus sombre la dure réalité pour un ARNm, statistiquement toujours en danger et constamment dans l'attente d'un début de dégradation. Ainsi les « jeunes » ARNm ont autant de chances d'être attaqués que les « vieux » ARNm. Quelques copies d'un ARNm peuvent subir plusieurs cycles de traduction, tandis que d'autres fonctionnent à peine. Cette espérance de vie aléatoire est caractéristique des molécules individuelles

d'ARNm procaryotes et eucaryotes. Mais le rendement traductionnel global de toute séquence de messenger est prévisible.

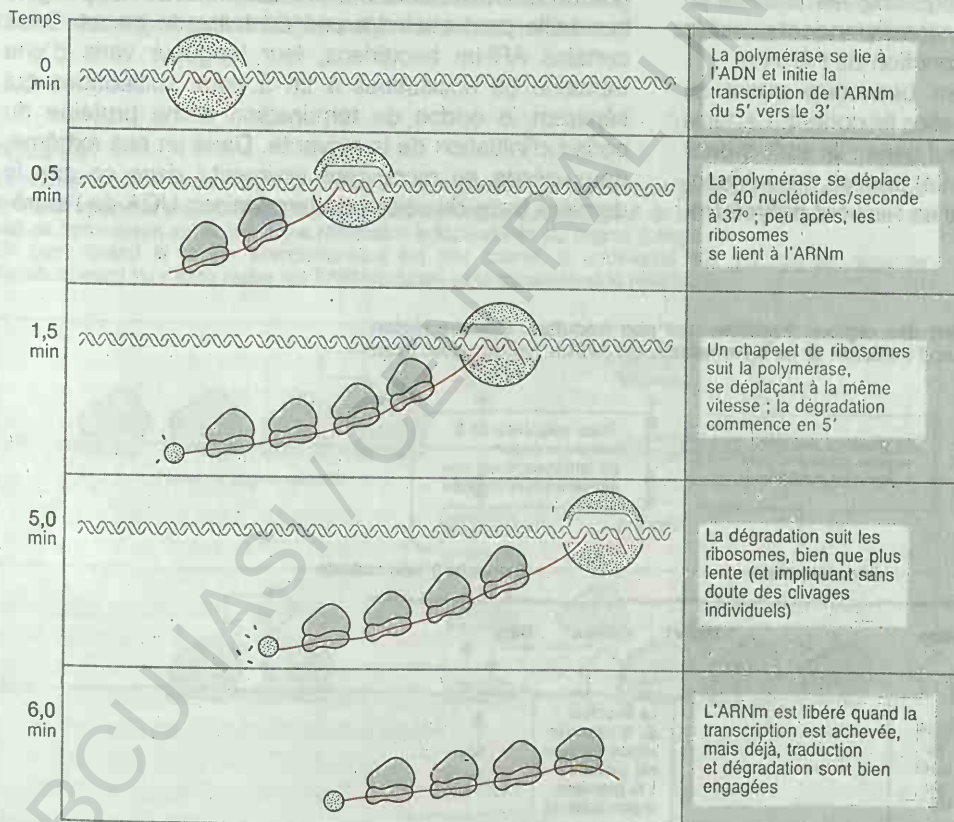
LA PLUPART DES ARNm BACTÉRIENS SONT POLYCISTRONIQUES

En raison de leur instabilité *in vivo*, les ARNm bactériens peuvent rarement être isolés intacts. Cependant, la région de l'ADN représentée dans l'ARNm peut être déterminée par des techniques d'hybridation, qui mesurent la capacité de l'ARNm marqué à s'hybrider aux fragments d'ADN qui représentent des parties spécifiques du gène. De cette façon, une image détaillée de la structure et de la

Figure 8.2

Transcription, traduction et dégradation s'effectuent simultanément dans la bactérie.

La figure montre une unité de transcription hypothétique d'environ 12 000 pb dont l'expression débute au temps 0. Pour simplifier, l'unité est représentée comme si elle était constituée d'un seul gène (une unité habituelle de cette longueur représenterait probablement plusieurs gènes).



fonction de l'ARNm bactérien peut être établie sans avoir à caractériser directement les molécules d'ARNm.

Les infections phagiques sont des exceptions utiles qui permettent l'obtention d'ARNm. Dans quelques cas, l'ARNm d'une cellule infectée par un phage est plus stable que celui d'un hôte bactérien et ceci permet la caractérisation physique de la molécule. L'exemple *par excellence* est fourni par l'ARN de phages, dans lesquels l'ARNm est partie intégrante du génome, d'où non seulement sa stabilité mais aussi la possibilité d'en obtenir de grandes quantités.

Dans certains cas, au moins, la stabilité accrue de l'ARNm phagique est une propriété intrinsèque du message. Le leader 5' d'un ARNm stable codé par le phage T4 confère une stabilité à l'ARNm bactérien auquel il est uni. Cette région résiste sans doute aux nucléases qui initient le mécanisme de dégradation de l'ARNm.

L'emploi de systèmes acellulaires pour la transcription et la traduction se révèle d'une grande utilité pour travailler directement sur l'ARNm bactérien. Un ARNm peut être transcrit *in vitro*, à partir d'une matrice appropriée, généralement une copie « clonée » du gène. Il peut être traduit par les ribosomes d'*E. coli*, les aminoacyl-ARNt, etc., en protéine. De tels systèmes rendent possible l'étude de la structure et de la fonction de l'ARNm.

Les ARNm bactériens varient beaucoup quant au nombre de protéines pour lesquelles ils codent. Certains ARNm ne représentent qu'un seul gène : ils sont **monocistroniques**. D'autres (la majorité) portent des séquences codant pour plusieurs protéines : ils sont **polycistroniques**.

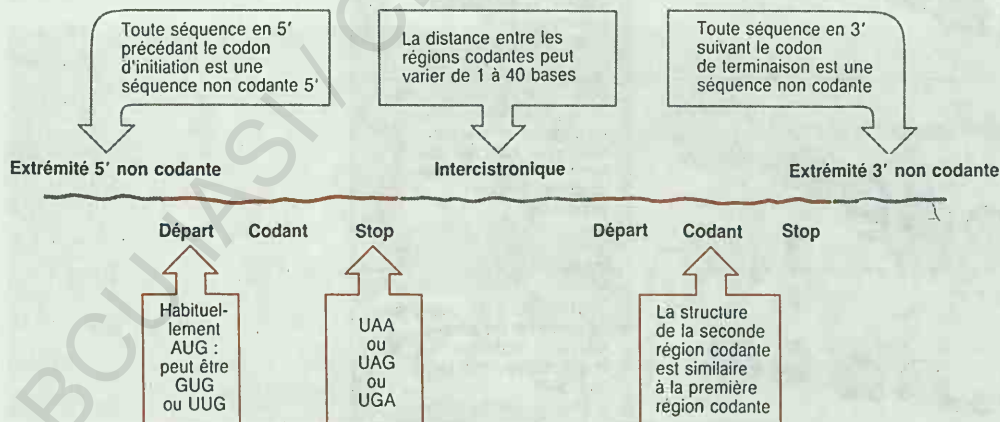
niques. Pour ceux-ci, un seul ARNm est transcrit à partir d'un groupe de gènes adjacents (comme nous le verrons dans le Chapitre 10, un tel groupement de gènes constitue un *opéron* qui est contrôlé comme une seule unité génétique).

Tous les ARNm contiennent deux types de régions. La **région codante** constituée d'une série de codons qui représentent la séquence en acides aminés de la protéine, débute (généralement) par AUG et s'achève par un codon de terminaison. Mais l'ARNm est toujours plus long que la partie codante. Dans un ARNm monocistronique, des régions supplémentaires peuvent apparaître aux extrémités. Précédant le début de la partie codante, une séquence additive en 5' est décrite comme **leader** (ou région 5' non codante). Une séquence supplémentaire qui suit le signal de terminaison en formant l'extrémité 3' est appelée « **trailer** » (ou extrémité 3' non codante). Bien qu'elles appartiennent à l'unité de transcription, ces séquences ne codent pas pour la protéine.

La **Figure 8.3** illustre la structure d'un ARNm polycistronique. Les **régions intercistroniques** qui s'étendent entre les différentes régions codantes présentent une grande variété de taille. Dans l'exemple d'ARN de phages, leur taille peut atteindre une centaine de bases. Dans certains ARNm bactériens, leur longueur varie d'une trentaine de nucléotides à un à deux nucléotides qui séparent le codon de terminaison d'une protéine du codon d'initiation de la suivante. Dans un cas extrême, deux gènes se recouvrent vraiment ; dans ce cas la dernière base du codon de terminaison UGA de l'extré-

Figure 8.3

L'ARNm bactérien contient aussi bien des régions traduites que non traduites. Chaque région codante possède ses propres signaux d'initiation et de terminaison. Un ARNm typique peut avoir plusieurs régions codantes.



mité d'une région codante est aussi la première base du codon d'initiation AUG qui commence un autre gène.

TRADUCTION DES MESSAGERS POLYCISTRONIQUES

Les vitesses auxquelles les chaînes d'ARN et de protéines s'allongent sont presque identiques pour tous les gènes, aussi la quantité de protéine synthétisée dépend d'abord de l'efficacité de l'initiation de la transcription et de la

Dans l'exemple des gènes d'*E. coli*, 15 initiations de transcription sont réalisées toutes les minutes. Chacun des 15 ARNm est sans doute traduit par trente ribosomes dans l'intervalle compris entre sa transcription et sa dégradation. Donc, à l'équilibre, 150 molécules de protéines sont traduites par minute.

Le nombre de ribosomes engagés dans la traduction d'un cistron particulier dépend de l'efficacité de son site d'initiation. Dès que l'extrémité 5' de l'ARNm est synthétisée, le site d'initiation du premier cistron devient disponible. Comment sont traduits les cistrons suivants ? Les différentes régions d'un ARNm polycistronique sont-elles

traduites séparément ou bien leur expression est-elle connectée ? Le mécanisme d'initiation est-il le même pour tous les cistrons, ou bien diffère-t-il entre le premier cistron et les cistrons internes ?

La traduction d'un ARNm bactérien progresse séquentiellement à travers ses cistrons. Alors que les ribosomes se lient à la première région codante, la transcription des régions codantes suivantes n'a même pas encore été effectuée. Au moment où le second site de fixation du ribosome est disponible au niveau du second cistron, la traduction a déjà bien avancé le long du premier cistron.

Ce qui se produit entre les régions codantes peut dépendre de chaque ARNm particulier. Sans doute dans presque tous les cas, les ribosomes s'attachent indépendamment au début de chaque cistron. La série de réactions la plus plausible apparaît dans la partie supérieure de la **Figure 8.4**. Une fois terminée la synthèse de la première protéine, les ribosomes se dissocient en sous-unités et quittent l'ARNm. Puis une nouvelle sous-unité 30S doit s'unir au codon d'initiation suivant, être rejointe par une sous-unité 50S, et s'organiser pour traduire le cistron suivant.

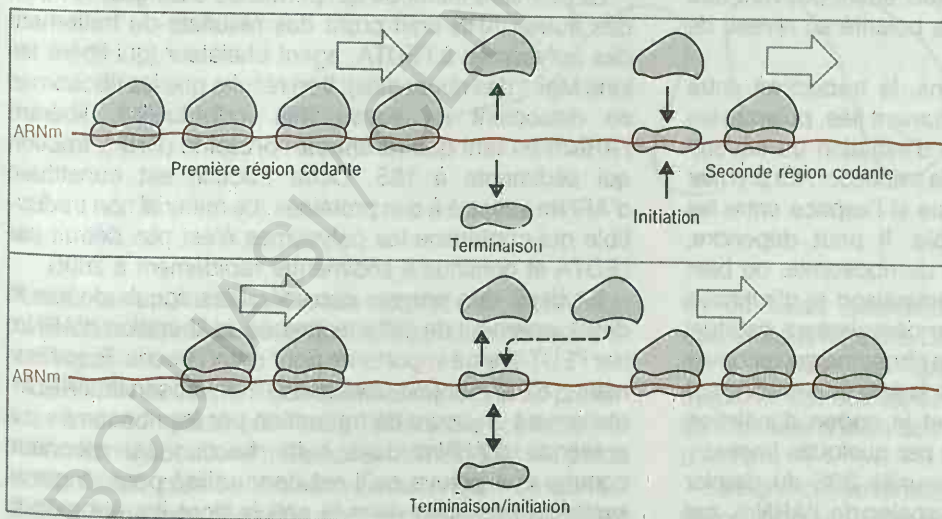
Mais cela ne signifie pas que la traduction d'un cistron est toujours sans effet sur la traduction du cistron suivant.

Figure 8.4

La réinitiation interne des ARNm polycistroniques peut être influencée par la région intercistronique.

En haut : quand la région intercistronique est plus longue que la largeur du ribosome, la dissociation au site de terminaison est suivie d'une réinitiation indépendante au cistron suivant.

En bas : quand la région intercistronique est très courte, la sous-unité 30S pourrait se dissocier transitoirement ou même rester sur l'ARNm durant la terminaison et la réinitiation.



Dans certaines unités, une mutation non-sens dans un gène inhibe l'expression d'un gène situé beaucoup plus loin sur l'ARNm polycistronique. Cet effet est désigné sous le terme de **polarité**. La cause principale de la polarité est un effet (indirect) sur la transcription ; la synthèse d'ARN cesse peu après le site muté (voir Chapitre 11). Cependant, la polarité peut aussi résulter d'une relation entre la traduction de deux cistrons dans l'ARNm.

La traduction d'un cistron peut nécessiter des changements de structure secondaire qui dépendent de la traduction du cistron précédent. Un effet de cette nature apparaît normalement dans la traduction de l'ARN phagique dont les cistrons sont toujours exprimés dans un ordre déterminé. La **Figure 8.5** montre que l'ARN phagique adopte une structure secondaire dans laquelle une seule séquence d'initiation est accessible ; la seconde ne peut pas être reconnue par les ribosomes car ses bases sont appariées à celles d'autres régions de l'ARN. Cependant, la traduction du premier cistron rompt la structure secondaire, permettant aux ribosomes de se lier au site d'initiation du cistron suivant. Dans cet ARNm, la structure secondaire contrôle la traductibilité.

Ce type de relation ne s'applique pas habituellement aux ARNm bactériens car les ribosomes suivent étroitement l'ARN polymérase, et les régions qui leur sont en amont n'ont pas la possibilité de s'apparier pour former des régions stables double brin. Mais, quand les ribosomes sont dans un cistron précoce, les régions ultérieures de l'ARNm peuvent former une structure secondaire. L'appariement de bases entre un site d'initiation et quelqu'autre région complémentaire pourrait prévenir la liaison de la sous-unité 30S. De tels effets peuvent être responsables de la création de la polarité au niveau de la traduction.

Chez certains ARNm bactériens, la traduction entre deux cistrons adjacents est absolument liée, puisque les ribosomes n'accèdent au codon d'initiation du second cistron que lorsqu'ils ont *achevé* la traduction du premier cistron. Ceci ne peut avoir lieu que si l'espace entre les deux régions codantes est faible. Il peut dépendre, localement, d'une haute densité de ribosomes ou bien de la juxtaposition de sites de terminaison et d'initiation qui permettrait de transgresser le déroulement habituel des réactions intercistroniques. Un ribosome qui recouvre 35 bases de l'ARNm peut donc à la fois être en contact avec un codon de terminaison et le codon d'initiation suivant s'ils ne sont séparés que par quelques bases.

Dans un cas extrême, la sous-unité 30S du dernier ribosome pourrait ne pas se dissocier de l'ARNm, car

elle est sur le champ attirée vers le site d'initiation dont elle est déjà virtuellement en possession. Comme l'illustre la moitié inférieure de la Figure 8.4, ceci signifierait que quand les sous-unités 50S et la chaîne polypeptidique achevées sont libérées, la sous-unité 30S resterait *in situ* pour initier à nouveau la traduction du cistron suivant.

Que le mécanisme de dépendance repose sur les changements de structure secondaire ou sur le transfert direct du ribosome d'un site de terminaison à un site d'initiation, il résulte d'une telle relation que la traduction de deux cistrons ou plus sur un ARNm polycistronique peut être contrôlée de façon coordonnée. On connaît des cas, chez les bactéries, où la non-traduction du premier cistron sur un ARNm empêche aussi la traduction des cistrons ultérieurs (voir Chapitre 10).

UNE DÉFINITION FONCTIONNELLE DE L'ARNm EUCARYOTE

Les polyribosomes eucaryotes sont généralement stables. Mais du fait que l'ARNm n'est qu'un constituant mineur de l'ARN total dans la fraction polysomale isolée (l'écrasante proportion est constituée par l'ARNr), il ne peut être isolé directement par les techniques de fractionnement habituelles. Des essais de marquage préférentiel en utilisant des précurseurs nucléotidiques radioactifs (approche utilisée pour les bactéries) n'ont pas été couronnés de succès, car certaines fractions d'ARN cytoplasmiques contaminantes sont marquées tout aussi rapidement et efficacement.

La première méthode qui permet de distinguer l'ARNm des autres ARN tirait profit des résultats du traitement des polysomes à l'EDTA, agent chélateur (qui libère les ions Mg^{2+} des ribosomes). Il en résulte que les ribosomes se dissocient en sous-unités individuelles, libérant l'ARNm en tant que **ribonucléoprotéine (RNP)**, fraction qui sédimente à 18S. Cette fraction est constituée d'ARNm associé à des protéines. Le matériel non traductible qui contamine les polysomes n'est pas détruit par l'EDTA et continue à sédimenter rapidement à 200S.

En dépit des progrès considérables acquis depuis le développement de cette technique, la libération d'ARNm par l'EDTA reste importante pour deux raisons. Premièrement, c'est un test *fonctionnel* : il reconnaît l'ARNm réellement en cours de traduction par les ribosomes. La présence d'ARNm dans cette fraction est reconnue comme une preuve qu'il est donc utilisé pour diriger la synthèse protéique dans la cellule dont il a été extrait.

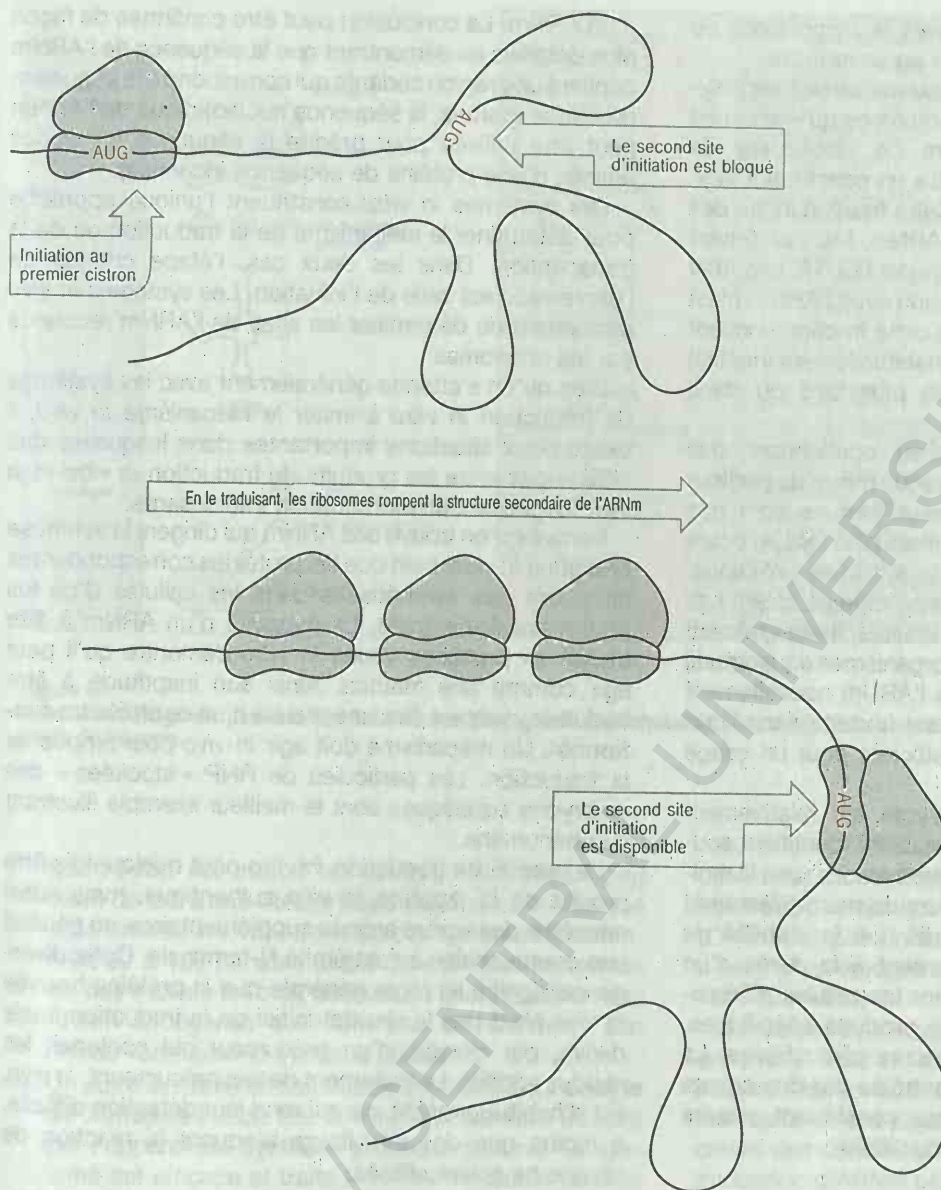


Figure 8.5
La structure secondaire peut contrôler l'initiation. Un seul site d'initiation est disponible dans l'ARN de phage, mais la traduction du premier cistron change la conformation de l'ARN, rendant ainsi disponible(s) d'autre(s) site(s) d'initiation.

Deuxièmement, cette technique permet d'isoler l'ARNm sous sa forme présumée naturelle, la particule ribonucléoprotéique.

La RNPm ne contient généralement que peu de protéines. Les comparaisons entre différentes cellules de mammifères aboutissent à deux protéines dont les poids moléculaires, plus ou moins constants, sont de 52 000 et 78 000 daltons. Peu d'autres protéines apparaissent

en de telles quantités ; la plupart ne sont que des constituants mineurs. Dans le cas le mieux étudié, l'ARNm de la globine des hématies, la quantité de protéines majeures est suffisante pour qu'une ou deux molécules de chacune d'elles soit associée avec chaque molécule d'ARNm.

La signification fonctionnelle de ces protéines n'est pas encore connue. Elles pourraient être impliquées dans le

transport de l'ARNm du noyau vers le cytoplasme ou pourraient influencer sa traduction ou sa stabilité.

La fraction ribonucléoprotéique qui contamine les polyosomes n'est en fait contaminante qu'en ce qu'elle est un constituant indésirable au cours de l'isolement de l'ARNm. La RNP est présumée être un constituant légitime à part entière de la cellule. Cette fraction inclut des molécules d'ARN semblables à l'ARNm. Mais en tenant compte de la méthode de relargage par l'EDTA, ces ARN ne sont pas traductibles. Leur relation avec l'ARNm n'est pas claire ; on peut envisager que cette fraction contient des ARNm légitimes qui (bien qu'habituellement inactifs) seront utilisés pour la traduction plus tard ou dans d'autres circonstances.

Certaines cellules embryonnaires contiennent des ARNm en réserve. On les retrouve sous forme de particules ribonucléoprotéiques dont les ARNm ne sont pas encore traduits, mais qui seront utilisés plus tard au cours de l'embryogenèse pour diriger la synthèse protéique. Bien que de tels mécanismes de stockage ne soient pas encore connus dans les cellules adultes, il est vraiment bien établi (en particulier chez les organismes aquatiques) qu'une partie non négligeable de l'ARNm nouvellement synthétisé n'est pas immédiatement traduite dans l'œuf ou l'embryon précoce mais est stockée pour un usage ultérieur.

L'ARNm du cytoplasme eucaryote est relativement stable. Les évaluations de cette stabilité identifient souvent plusieurs composants. Typiquement, environ la moitié des ARNm de cellules en culture de mammifères ont une demi-vie d'environ 6 h, tandis que la stabilité de l'autre moitié équivaut grossièrement à la durée d'un cycle cellulaire, environ 24 h *. Dans les cellules différenciées destinées à la synthèse de produits spécifiques, quelques ARNm peuvent être encore plus stables. La stabilité d'un ARNm peut être contrôlée par des agents externes, par exemple les hormones, constituant un autre mécanisme de contrôle de la traductibilité.

LA PUISSANCE DES SYSTÈMES DE TRADUCTION *IN VITRO*

La possibilité de traduire l'ARNm *in vitro* apporte un outil crucial dans la détermination du procédé de la traduction. Montrer que le produit de la traduction possède les propriétés authentiques d'une protéine trouvée *in vivo* apporte l'évidence sans équivoque de la fonction codante

d'un ARNm. La conclusion peut être confirmée de façon plus détaillée en démontrant que la séquence de l'ARNm contient une région codante qui correspond à la séquence protéique (parfois, la séquence nucléotidique de l'ARNm peut être utilisée pour prédire la séquence en acides aminés d'une protéine de séquence inconnue).

Les systèmes *in vitro* constituent l'unique approche pour déterminer le mécanisme de la traduction ou de la transcription. Dans les deux cas, l'étape critique de l'expression est celle de l'initiation. Les systèmes *in vivo* permettent de déterminer les sites de l'ARNm reconnus par les ribosomes.

Bien qu'on s'attende généralement avec les systèmes de traduction *in vitro* à imiter le mécanisme *in vivo*, il existe deux situations importantes dans lesquelles des différences entre les produits de traduction *in vitro* et *in vivo* fournissent une information intéressante.

Il arrive qu'on trouve des ARNm qui dirigent la synthèse protéique *in vitro*, bien que les protéines correspondantes ne soient pas synthétisées dans les cellules d'où les ARNm ont été extraits. La capacité d'un ARNm à être traduit en protéines vraies *in vitro* démontre qu'il peut agir comme une matrice. Ainsi son inaptitude à être traduit *in vivo* peut être une preuve d'un **contrôle traductionnel**. Un mécanisme doit agir *in vivo* pour empêcher la traduction. Les particules de RNP « stockées » des embryons aquatiques sont le meilleur exemple illustrant ce phénomène.

Le produit de traduction *in vitro* peut quelquefois être proche de la protéine *in vivo* authentique, mais aussi posséder des acides aminés supplémentaires, en général une courte chaîne à l'extrémité N-terminale. Cette divergence signifie en règle générale que la protéine trouvée *in vivo* n'est pas le résultat initial de la traduction, mais dérive, par clivage, d'un précurseur qui contenait les résidus additifs. Le traitement de tels précurseurs, *in vivo*, est d'habitude rapide, ce qui rend leur détection difficile, à moins que des inhibiteurs bloquant la réaction de clivage ne soient utilisés.

Les ARNm isolés ont été caractérisés par leur traduction dans deux types de système, décrits dans la Figure 8.6.

- Les systèmes de synthèse acellulaire renferment les ribosomes, les facteurs de synthèse protéique et les ARNt. Un grand nombre de systèmes de ce type ont été développés ; les plus connus sont dérivés d'*E. coli*, de germe de blé et de réticulocytes de lapin. Les systèmes sont quelque peu inefficaces, puisque cha-

* N. d. T. : Certains messagers eucaryotiques, notamment ceux codant pour des protéines intervenant dans le cycle cellulaire, ou pour des produits d'oncogènes, ont des demi-vies de l'ordre de quelques minutes.

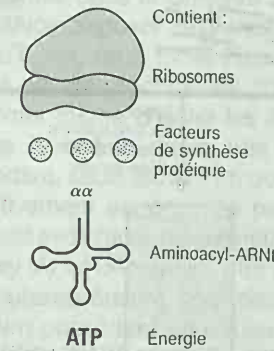
		Avantages	Inconvénients
Systèmes acellulaires	 Contient : Ribosomes Facteurs de synthèse protéique Aminoacyl-ARNt ATP Énergie	Peut être utilisé pour tester le rôle de chaque constituant Composé protéique facile à purifier Peut permettre d'obtenir des protéines précurseurs	Opère lentement et inefficacement (par rapport au taux de synthèse protéique <i>in vivo</i>) Les conditions de la réaction diffèrent de celle <i>in vivo</i> (telles que les concentrations des constituants)
Oeuf de Xénope	 Injection de l'ARNm à la micropipette Oocyte	Très efficace, opère de la même façon qu'<i>in vivo</i>	Les produits protéiques sont plus difficiles à purifier

Figure 8.6

Les ARNm exogènes peuvent être traduits par des systèmes acellulaires ou par injection dans les œufs de Xénope.

que ARNm est traduit un nombre relativement peu élevé de fois à une vitesse beaucoup moins élevée que celle trouvée *in vivo*. Au mieux, ces systèmes fonctionnent 90 à 120 minutes avant de s'arrêter. Dans tous les cas il existe un bruit de fond dû à la traduction des ARNm endogènes qui n'ont pas été éliminés ; son niveau varie avec le système utilisé.

- Un autre système de traduction est fourni par l'**oocyte de Xénope** intact. Les ARNm injectés sont traduits par l'appareil de synthèse protéique naturel. Le système est efficace et traite les ARNm injectés comme s'ils étaient des ARNm endogènes, ce qui leur permet d'être utilisés pour plusieurs cycles de traduction. La seule limite est qu'une trop grande quantité d'ARNm peut saturer le système traductionnel (qui est présent en large excès par rapport à l'offre que constituent les ARNm endogènes). En général, le système continue à être actif pendant 24 à 48 heures.

Aucun de ces systèmes ne présente de spécificité vis-à-vis du tissu ou de l'espèce. Ceci indique que les ARNm et l'appareil de synthèse protéique de (peut-être) tous les cytoplasmes sont interchangeables. Il ne semble

pas y avoir de contrôle traductionnel comme des facteurs ou des ribosomes qui fonctionneraient avec un ensemble d'ARNm et pas un autre. La machinerie traductionnelle n'est pas préprogrammée, mais acceptera tout ARNm comme matrice.

Le contrôle traductionnel doit prévenir l'interaction entre les ARNm et l'appareil de synthèse protéique. Ceci peut se réaliser soit en les séquestrant sous une forme telle qu'ils soient physiquement non disponibles, soit en créant des conditions de compétition dans lesquelles la traduction d'ARNm plus efficaces s'effectue aux dépens d'ARNm qui le sont moins.

Le système oocyte peut permettre la maturation d'au moins quelques protéines qui sont normalement clivées au cours de leur synthèse ou peu après. Dans certains cas, il peut même diriger l'entrée des protéines dans des fac-similés du compartiment cellulaire approprié. Cette capacité implique que des signaux de traitement peuvent être communs à différents types de cellules et d'espèces. En conséquence pratique, ceci signifie que l'isolement de précurseurs protéiques peut être entrepris dans des systèmes acellulaires.

LA PLUPART DES ARNm EUCARYOTES SONT POLYADÉNYLÉS EN 3'

La plupart des ARNm ont une séquence polyadénylée à l'extrémité 3'. Cette extension terminale de résidus A est souvent décrite comme la **queue poly (A)** et l'ARNm qui possède cette caractéristique est noté : **poly (A)⁺**.

La séquence poly (A) n'est pas codée dans l'ADN, mais l'addition de poly (A) est catalysée par la poly (A) polymérase, qui ajoute environ 200 résidus A à l'extrémité 3' – OH libre de l'ARNm. Nous ne savons pas comment s'effectue le contrôle de la longueur de l'extension.

Quand l'ARNm pénètre dans le cytoplasme, la queue poly (A) est approximativement de la même longueur que celle ajoutée dans le noyau. Elle est graduellement écourtée, peut-être au cours d'étapes successives impliquant un clivage endonucléolytique. La population cellulaire d'ARNm est constituée à la fois de « nouvelles » et d'« anciennes » molécules, avec des extensions poly (A) respectivement plus longues ou plus courtes. Mais, la longueur de la queue poly (A) présente sur chaque molécule ne semble pas exercer d'influence sur sa capacité à être traduite ou sa stabilité dans le cytoplasme.

Le poly (A) d'ARNm de mammifère est associé avec la protéine commune de 78 000 daltons, constituant majeur de la RNPM. Ceci suggère que la structure de l'extrémité 3' de l'ARNm consiste en une extension de poly (A) liée à une masse de protéines grossièrement identiques. Cette localisation de la protéine explique au moins pourquoi ce constituant de la RNPM n'interfère pas avec la traduction.

La présence de poly (A) a une conséquence pratique importante. La partie poly (A) d'un ARNm peut se lier par appariement de bases à l'oligo (U) ou l'oligo (dT) ; cette réaction peut être utilisée pour isoler les ARNm poly (A)⁺. La technique la plus commode consiste à immobiliser l'oligo (U ou dT) sur un support solide. Puis quand une population d'ARN est passée sur la colonne, comme l'illustre la **Figure 8.7**, seul l'ARN poly (A)⁺ est retenu. Il peut être récupéré en traitant la colonne avec une solution qui rompt les liaisons et libère l'ARN.

Le seul inconvénient de cette méthode est qu'elle isole tous les ARNm polyadénylés. Si par exemple l'ARN de la cellule entière est employé, les ARNm poly (A)⁺ nucléaires et cytoplasmiques seront retenus. Si des préparations de polysomes sont utilisées (un procédé courant), presque tous les ARNm poly (A)⁺ isolés seront des ARNm actifs ; mais quelques ARNm de la fraction RNP contaminante, également polyadénylés, seront aussi isolés. Ceci rend

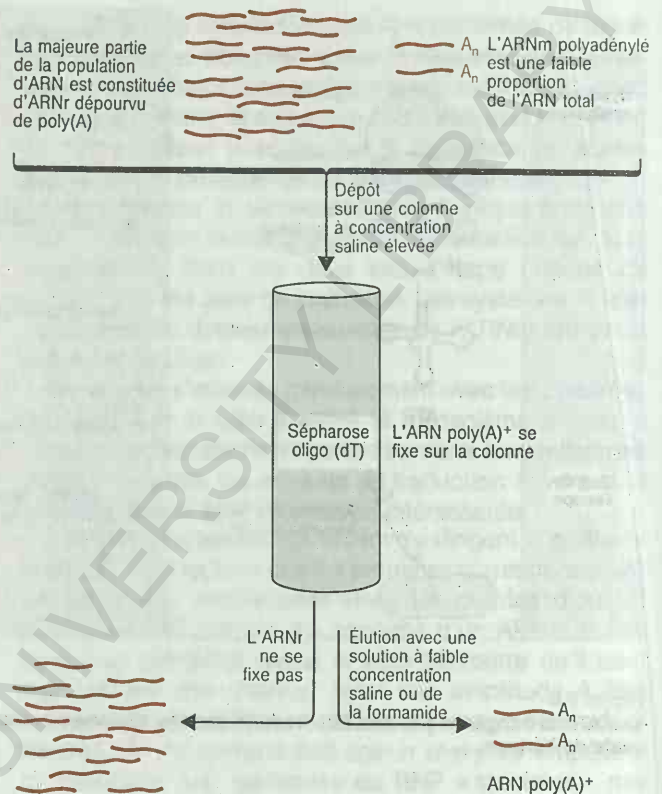


Figure 8.7
Les ARNm poly(A)⁺ peuvent être séparés des autres ARNm par fractionnement sur Sépharose-oligo(dT).

nécessaire l'utilisation du dosage de relargage à l'EDTA comme test fonctionnel quand il est important d'isoler précisément la population active d'ARNm (certains oublient souvent cette opposition et traitent l'ARN poly (A)⁺ cytoplasmique comme s'il était l'équivalent de l'ARNm).

Le clonage est une approche de purification de l'ARNm qui utilise aussi la queue poly (A). D'abord, un petit fragment d'oligo (dT) est hybridé à l'extension poly (A) fournissant un point de départ ou amorce (primer) pour la **transcriptase inverse**. Cette enzyme copie l'ARNm pour élaborer un brin d'ADN complémentaire (connu en tant que **ADNc**). Puis l'ADNc peut servir de matrice pour synthétiser un brin d'ADN qui est identique à la séquence d'origine de l'ARNm. Le produit de ces réactions est un

ADN bicaténaire qui correspond à la séquence de l'ARNm. Cet ADN peut être rapidement cloné par les techniques décrites dans le Chapitre 16.

Le fait de pouvoir disposer de grandes quantités d'ADN cloné, quel qu'il soit, rend facile l'isolement de l'ARNm correspondant par des techniques d'hybridation. Cette technique permet même d'isoler les ARNm qui ne sont présents dans la cellule qu'en très petit nombre de copies. Cependant, seuls les ARNm qui sont présents en quantités relativement importantes peuvent être isolés directement sans avoir recours aux méthodes de clonage.

Quand le poly (A) fut découvert, il fut envisagé que tous les ARNm cellulaires étaient polyadénylés. Bien que la fraction d'ARNm possédant une queue poly (A) ait toujours été inférieure à 100 %, cette différence était attribuée à une dégradation au cours de la préparation (une seule rupture dans l'ARNm poly (A)⁺ engendrerait une extrémité 5' sans poly (A) ; ce fragment pourrait être confondu avec un ARNm vrai). Cependant, même avec des techniques de préparation de l'ARNm extrêmement soignées, dans lesquelles la dégradation était réduite au minimum, une fraction d'ARNm était toujours dépourvue de poly (A). Ceci représente jusqu'à un tiers de l'ARNm total.

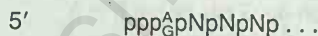
Un constituant prédominant de cette fraction **poly (A)⁻** est fourni par les ARNm qui codent pour les protéines histones des chromosomes. La partie restante de la fraction poly (A)⁻ (environ les deux tiers) est identique à la fraction poly (A)⁺ à tout point de vue excepté la présence de poly (A). Longueur, stabilité, efficacité de traduction et transport nucléocytoplasmique semblent être les mêmes pour les ARNm poly (A)⁻ et poly (A)⁺ (la présence de poly (A) ne semble pas augmenter la stabilité de certains ARNm quand ils sont injectés dans l'oocyte de *Xenopus* ; mais l'effet n'est ni général, ni apparemment applicable à la cellule normale).

Les ARNm poly (A)⁺ et poly (A)⁻ codent-ils ou non pour les mêmes protéines ? Les ARNm de l'histone sont spécifiques de la fraction poly (A)⁻, mais les autres constituants de la fraction poly (A)⁻ présentent un chevauchement considérable avec ceux de la fraction poly (A)⁺. Tous ces ARNm peuvent aussi exister sous forme polyadénylée. Ainsi, un gène peut être représenté par des transcrits dont seuls certains sont poly (A) et d'autres non. Il peut exister des différences dans le pourcentage de polyadénylation de chaque ARNm, mais aujourd'hui encore, plus de dix ans après sa découverte, nous n'avons aucune idée de la signification de la polyadénylation (ou de son absence).

TOUS LES ARNm EUKARYOTES ONT UNE COIFFE (CAP) MÉTHYLÉE À L'EXTRÉMITÉ 5'

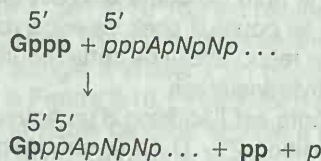
L'extrémité 5' de l'ARNm est modifiée dans le cytoplasme eucaryote (contrairement à la mitochondrie ou au chloroplaste). Les réactions de modification sont sans doute communes à tous les eucaryotes.

La transcription commence avec un nucléotide triphosphate (d'habitude une purine A ou G). Le premier nucléotide conserve son groupe 5' triphosphate et forme la liaison phosphodiester courante entre sa position 3' et la position 5' du nucléotide suivant. La séquence initiale du transcrit peut être représentée ainsi :



Mais quand l'ARNm est traité *in vitro* par des enzymes qui devraient le dégrader en nucléotides individuels, l'extrémité 5' ne fait pas apparaître le nucléoside triphosphate attendu. Elle contient au contraire, deux nucléotides, reliés par un *pont 5'-5' triphosphate* et porte aussi des groupements méthyle. La base terminale, toujours une guanine, vient s'ajouter à l'ARN d'origine *après la transcription*.

La guanylyl transférase, enzyme nucléaire, catalyse l'addition de G terminal en 5'. Cette réaction s'effectue tellement rapidement après le début de la transcription qu'il n'est possible de détecter que des traces infimes de l'extrémité 5' triphosphate d'origine dans l'ARN nucléaire. La réaction globale peut être représentée comme une condensation entre le GTP et l'extrémité 5' triphosphate d'origine de l'ARN. Donc



Le nouveau résidu **G** ajouté à l'extrémité de l'ARN est orienté à l'inverse de tous les autres nucléotides.

Cette structure porte le nom de **coiffe** (cap en anglais). Elle sert de substrat à plusieurs méthylations. La **Figure 8.8.** montre la structure complète de la coiffe après addition de tous les groupements méthyle possible.

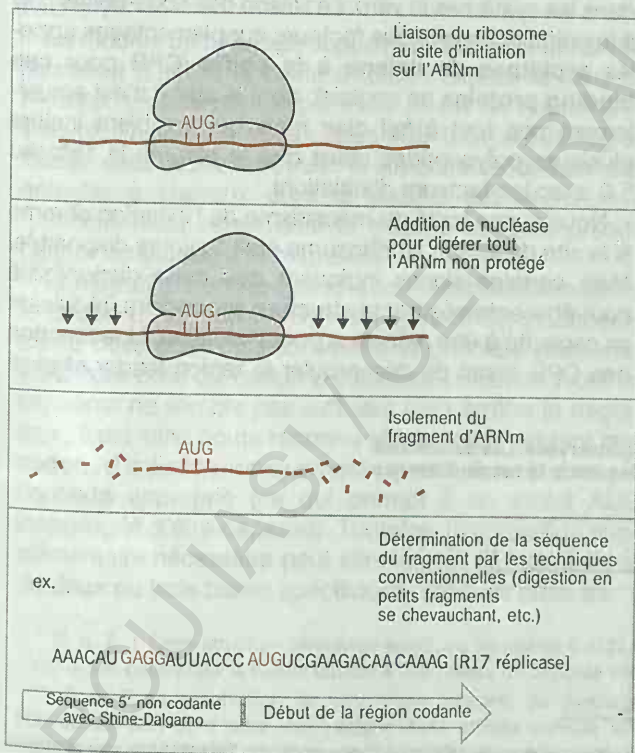
La première méthylation est présente chez tous les eucaryotes et consiste en l'addition d'un groupement méthyle sur la position 7 de la guanine terminale. Une coiffe qui possède cet unique groupement méthyle est

rieurs. Ceci s'accomplit par la production de résidus N⁶ méthyladénine à la fréquence d'environ une modification pour mille bases. Il peut y avoir 1 à 2 méthyladénines par ARNm d'un eucaryote supérieur typique, bien que leur présence ne soit pas obligatoire puisque certains ARNm (par exemple la globine) n'en ont aucune.

L'INITIATION PEUT IMPLIQUER L'APPARIEMENT DE BASES ENTRE L'ARNm ET L'ARNr

Les sites d'initiation de la synthèse protéique peuvent être identifiés par liaison des ribosomes à l'ARNm dans des conditions qui bloquent l'élongation. Le ribosome reste alors au site d'initiation. Quand le complexe d'initiation bloqué est mis en présence de ribonucléase, les régions de l'ARNm en dehors du ribosome sont dégradées, alors que celles qui lui sont réellement liées sont

Figure 8.9
Les sites de liaison des ribosomes à l'ARNm peuvent être récupérés à partir des complexes d'initiation.



protégées ainsi que l'illustre la **Figure 8.9**. Les fragments protégés peuvent être récupérés et caractérisés.

Les séquences d'initiation protégées par les ribosomes bactériens ont une longueur d'environ 35 à 40 bases. Il existe très peu d'homologie entre les sites de liaison des ribosomes des différents ARNm bactériens. Ils n'ont que deux caractères en commun. Premièrement, le codon d'initiation AUG (ou plus rarement GUG ou UUG) fait toujours partie de la séquence protégée. Deuxièmement, ils contiennent une courte séquence qui est complémentaire d'une partie d'un hexamère situé près de l'extrémité 3' de l'ARNr 16S.

Écrit en sens inverse, l'hexamère devient :

3' ... U C C U C C ... 5'

Tous les sites d'initiation de l'ARNm d'*E. coli*, sauf un, contiennent une séquence complémentaire d'une partie nucléotidique au moins de cet hexamère et plus couramment complémentaire de 4-5 bases. Ainsi, l'ARNm bactérien contient une partie si ce n'est la totalité de l'oligonucléotide :

5' ... A G G A G G ... 3'

Cette séquence polypurinique est connue sous le nom de séquence de **Shine-Dalgarno**. Elle se situe 7 bases avant le codon AUG.

Cette séquence de Shine-Dalgarno s'apparie-t-elle à son complément sur l'ARNr au cours de la liaison ARNm-ribosome ? Un changement dans la séquence peut bloquer la traduction. Par exemple, la liaison du ribosome à l'ARNr du gène 0,3 du phage T7 est déstabilisée par une mutation qui transforme la séquence GAGG en GAAG.

L'extrémité 3' de l'ARNr 16S est hautement conservée chez les bactéries. Elle est auto-complémentaire et pourrait former l'épingle à cheveux en bases appariées qui apparaît sur la **Figure 8.10**. La séquence complémentaire de l'ARNm fait partie de cette épingle à cheveux potentielle comme souligné dans la figure.

La séquence complémentaire dans l'ARNr ne peut simultanément faire partie de l'épingle à cheveux (intra-moléculaire) et se lier à l'ARNm. Ces associations pourraient être alternatives, auquel cas la réaction d'initiation impliquerait la rupture de l'épingle permettant l'appariement de bases entre l'ARNm et l'ARNr. Après l'initiation, le duplex ARNm-ARNr pourrait être rompu par reconstitution de l'épingle à cheveux de l'ARNr en bases appariées. Ce mécanisme pourrait concilier la nécessité de former un complexe d'initiation stable en un site particulier avec la nécessité plus tardive de migrer le long de l'ARNm.

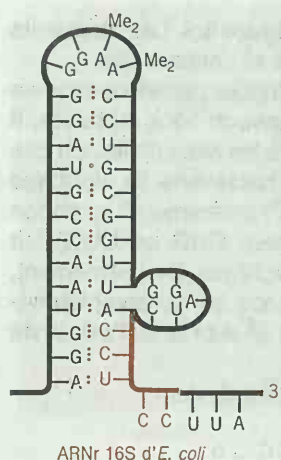


Figure 8.10

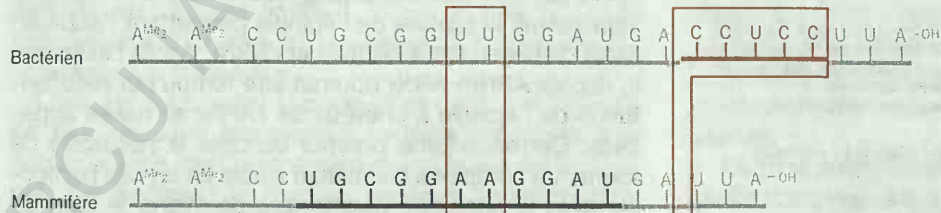
Les épingles à cheveux peuvent se former près de l'extrémité 3' de l'ARNr : elles contiennent les régions qui pourraient être impliquées dans la reconnaissance de l'ARNm bactérien (indiquées en rouge).

La Figure 8.11 montre que la séquence de l'extrémité 3' de l'ARNr est bien conservée entre les procaryotes et les eucaryotes. Il n'y a que deux changements significatifs dans les 20 nucléotides situés entre les adénines doublement méthylées adjacentes et l'extrémité 3'. Chez les bactéries, il y a deux résidus U, alors que chez les eucaryotes supérieurs il y a deux résidus A (chez les eucaryotes inférieurs, la séquence est un intermédiaire (AU). Et chez tous les eucaryotes, il y a une délétion d'une séquence de 5 bases CCUCC qui est le complément principal de la séquence de Shine-Dalgarno.

Plus d'une fois, on a suggéré qu'une autre séquence de l'ARNr eucaryote pourrait être une contrepartie du consensus procaryote de Shine-Dalgarno. Cependant, il n'existe aucune preuve pour étayer l'hypothèse d'un appariement de bases entre l'ARNm eucaryote et l'ARNr 18S.

Figure 8.11

Les extrémités 3' des petits ARNr bactériens et de mammifères sont bien conservées. Les différences sont indiquées dans les boîtes ; les régions qui peuvent être impliquées dans la reconnaissance de l'ARNm bactérien sont indiquées en rouge.



* N. d. T. : En réalité, plusieurs éléments de contrôle traductionnel sont localisés dans cette extrémité 3' non codante. Tel est le cas, par exemple, du motif d'ARN contrôlant la stimulation par le fer de la traduction des 3 messagers de la ferritine et du récepteur de la transferrine.

LES PETITES SOUS-UNITÉS PEUVENT MIGRER VERS LES SITES D'INITIATION SUR L'ARNm EUCARYOTE

En théorie, tous les ARNm eucaryotes sont monocistroniques, mais la séquence de chaque ARNm est habituellement plus longue que celle nécessaire au codage de sa protéine. La région codante n'occupe donc qu'une partie du message. Dans le cytoplasme eucaryote l'ARNm moyen est long de 1 000 à 2 000 paires de bases ; il a une coiffe méthylée à son extrémité 5' et porte 100 à 200 bases de poly (A) à son extrémité 3'.

Le « leader » non traduit en 5' est relativement court, constitué d'habitude (mais pas toujours) de moins d'une centaine de bases. La longueur de la région codante est fonction de la taille de la protéine. L'extrémité 3' non codante est souvent plus longue, elle comprend quelquefois un millier de bases. En raison de sa localisation, le leader ne peut pas être ignoré durant l'initiation, mais aucune fonction ne peut encore être attribuée à l'extrémité 3' non codante dans la traduction*.

Les ribosomes du cytoplasme eucaryote ne s'unissent pas directement au site d'initiation au début de la région codante. En fait, la coiffe méthylée à l'extrémité 5' est la première structure reconnue. Les messagers dont les coiffes ont été éliminées ne sont pas traduits efficacement dans les systèmes *in vitro*. La liaison des sous-unités 40S à la coiffe nécessite des facteurs supplémentaires appelés **protéines de liaison à la coiffe** (CPB pour **cap binding proteins** en anglais), dont le statut n'est actuellement pas tout à fait clair mais qui semblent inclure plusieurs polypeptides (ainsi que le résume le Tableau 5.3 avec les facteurs d'initiation).

Nous avons traité du mécanisme de l'initiation comme si le site de liaison au ribosome était toujours disponible. Mais certains signes indiquent que cette disponibilité peut être entravée par la structure secondaire, réduisant sa capacité à être reconnue pour l'initiation. Une fonction des CPB serait de désenrouler la région leader et ainsi

d'assurer la disponibilité des sites de liaison au ribosome sous forme simple brin.

La modification de l'extrémité 5' intervient dans presque tous les ARNm cellulaires et viraux ; elle est essentielle pour la traduction dans le cytoplasme eucaryote (bien qu'elle ne soit pas nécessaire pour les organelles). L'unique exception à cette règle est fournie par quelques ARNm viraux (tel que le poliovirus) qui n'ont pas de coiffe ; seuls ces ARNm viraux exceptionnels peuvent être traduits sans coiffe *in vitro* ; ils doivent donc posséder une autre structure alternative qui rend la coiffe inutile*.

Certains virus tirent profit de cette différence. L'infection par le poliovirus inhibe la traduction des ARNm de l'hôte. Ceci peut résulter d'une interférence avec les CPB, nécessaires à l'initiation des ARNm cellulaires, mais superflues pour l'ARNm de poliovirus sans coiffe.

Quelquefois, le codon, d'initiation AUG se situe dans les 40 bases de l'extrémité 5' de l'ARNm, de telle sorte que la coiffe et AUG sont recouverts par le ribosome. Mais dans beaucoup d'ARNm, la coiffe et AUG sont éloignés l'un de l'autre, environ 1 000 bases les séparent dans les cas extrêmes. Malgré tout, la présence de la coiffe est toujours nécessaire pour qu'un complexe stable se forme au codon d'initiation. Comment le ribosome peut-il se repérer sur deux sites aussi éloignés l'un de l'autre ?

Le modèle du balayage suppose que la sous-unité 40S reconnaît d'abord la coiffe 5' puis « migre » le long de l'ARNm. Le balayage à partir du 5' est un mécanisme linéaire. Quand les sous-unités 40S balayent la région leader, elles peuvent détruire la structure secondaire des épingles à cheveux dont la stabilité est inférieure à -30kcal, mais celles dont la stabilité est plus grande gênent ou bloquent la migration.

La migration s'arrête quand la sous-unité 40S rencontre le codon d'initiation AUG (voir Figure 5.16). Habituellement, mais pas obligatoirement, la première séquence AUG sera celle du codon d'initiation. Le codon AUG en lui-même ne semble pas suffisant pour arrêter la migration ; il est sans doute reconnu efficacement en tant que codon d'initiation uniquement quand il est dans le contexte approprié (ce qui permet à un triplet AUG inapproprié d'être dépassé). Toutefois l'information supplémentaire nécessaire peut simplement être constituée de deux ou trois bases spécifiques additives proches**.

* N. d. T. : Cette structure alternative serait une séquence d'ARN jouant le même rôle que la séquence de Shine-Dalgarno chez les procaryotes, c'est-à-dire permettant la liaison directe à son niveau de la petite sous-unité ribosomale.

** N. d. T. : L'observation de plus d'une centaine de messagers eucaryotiques a permis de décrire la séquence la plus fréquemment trouvée de part et d'autre des codons AUG utilisés comme initiateurs : Δ^{CC} AUG G. La purine en position -3 du codon d'initiation (adénine dans 80 % des cas et guanine dans 20 %) semble particulièrement importante pour qu'un codon méthionine soit utilisé comme codon d'initiation.

Quand la séquence leader est longue, une seconde sous-unité 40S pourrait reconnaître l'extrémité 5' avant que la première n'ait quitté le site d'initiation ; il pourrait donc y avoir une série de sous-unités progressant du leader vers le site d'initiation.

La liaison se stabilise au site d'initiation. Quand la sous-unité 40S est rejointe par la sous-unité 60S le ribosome est localisé en un site identifiable par le test de protection. Une sous-unité 40S protège une région d'une soixantaine de bases ; une fois les sous-unités 60S unies au complexe, la région protégée diminue jusqu'à atteindre la même longueur que chez les procaryotes, soit 30 à 40 bases (La réduction de la longueur protégée par les ribosomes 80S pourrait être due à un changement de conformation déclenchée dans la petite sous-unité du fait de son association avec la grande sous-unité ou pourrait représenter la perte des facteurs d'initiation qui protègent directement une région supplémentaire de l'ARNm).

L'idée que les ribosomes doivent commencer à l'extrémité 5' est compatible avec la nature monocistronique des ARNm eucaryotes. Certains ARNm viraux, rares, contiennent plus d'une région codante. Dans ces cas-là, seule la région codante la plus proche de l'extrémité 5' est traduite à partir de l'ARNm intact. Les autres ne peuvent être lues que si l'ARNm a été clivé pour engendrer une nouvelle extrémité 5' à proximité du codon interne d'initiation. Ce mécanisme renforce l'hypothèse que les sites d'initiation internes ne sont pas reconnus directement, à l'exception peut-être du cas des ARN viraux dépourvus de coiffe*.

LA SYNTHÈSE PROTÉIQUE EST LIÉE À LA LOCALISATION CELLULAIRE

La cellule eucaryote est une structure bien ordonnée dont les fonctions s'exercent à des emplacements spécifiques. Cette idée générale s'applique en particulier à la synthèse protéique. On peut diviser les polyribosomes en deux classes : « libres » et « liés à la membrane ». Ils sont engagés dans la synthèse de différents groupes de protéines. Les polysomes « libres » synthétisent les protéines qui n'interagissent pas avec les membranes. Les polysomes liés à la membrane synthétisent des protéines dont les localisations spécifiques à l'intérieur de la cellule dépendent de leur

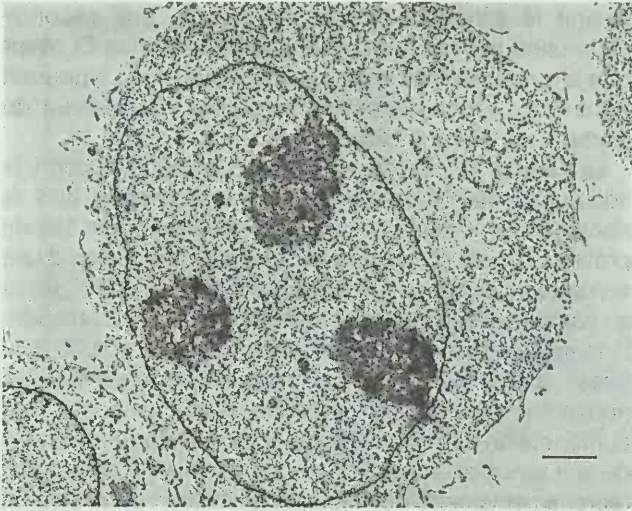


Figure 8.12
Les polyribosomes sont associés au réseau fibreux du cytosquelette.

Quand les cellules sont extraites avec un détergent non ionique (Triton) en solution tampon hypertonique, la plupart des lipides et des protéines sont éliminés, laissant une charpente cytosquelettique ancrée au noyau restant.

capacité à se lier ou à passer à travers les membranes (une exception étant la membrane mitochondriale).

Le qualificatif « libre » attribué aux polysomes est quelque peu inapproprié puisqu'en fait ils ne peuvent diffuser librement à travers le cytoplasme mais sont associés au **cytosquelette**, une ossature constituée d'un réseau complexe de fibres. Tous les polysomes libres sont associés au cytosquelette, comme le montre la photographie en microscopie électronique de la **Figure 8.12**.

Les polysomes ont tendance à s'agglomérer près du noyau, là où les ARNm pénètrent dans le cytoplasme. La plupart des produits de traduction sont des protéines solubles qui diffusent rapidement loin de leur site de synthèse quand elles sont libérées des polysomes. Les protéines constitutives du cytosquelette sont plutôt, incorporées en des endroits proches de leur site de synthèse.

Contrairement aux polysomes, les ribosomes monomériques sont totalement libres dans la cellule. L'association au cytosquelette est une fonction qui peut appartenir à l'ARNm. La traduction de l'ARNm peut être reliée à son association au cytosquelette, comme l'indiquent les événements qui interviennent dans l'infection virale. Dans l'exemple de l'infection par le VSV (vesicular stomatitis virus), virus de la stomatite vésiculaire, les ARNm viraux s'associent au cytosquelette peu après leur synthèse et

ne sont traduits que pendant cette période. Puis, ils se libèrent du cytosquelette.

L'existence de systèmes *in vitro* de traduction d'ARNm montre que la propriété de synthétiser des protéines peut être dégagée des structures cellulaires étrangères. Pour autant cela ne diminue pas l'importance de l'observation que le déroulement normal des événements pour l'ARNm est d'être associé au cytosquelette pendant la traduction. Ceci peut être important dans le métabolisme de l'ARNm *in vivo* de même que la nécessité d'une localisation particulière des protéines du cytosquelette nouvellement synthétisées.

Les protéines synthétisées par les polysomes liés à la membrane ont plusieurs destinations. Certaines sont séquestrées au sein de compartiments tels que le lysosome. Parmi celles qui constituent la membrane, l'attention la plus grande a été portée à celles qui résident dans la membrane plasmique (qui limite le cytoplasme). D'autres protéines seront **secrétées** de la cellule vers le milieu extracellulaire. Comment ces protéines trouvent-elles leurs sites terminaux ?

Pour beaucoup de protéines qui doivent être incorporées dans les membranes, la séquence du polypeptide mature n'est pas en elle-même suffisante pour diriger l'insertion dans la membrane. Une information supplémentaire est nécessaire ; elle se présente généralement sous la forme d'une **séquence leader** ou peptide signal en N-terminal de la protéine. Une **préprotéine** est une protéine qui porte ce leader. C'est un précurseur transitoire de la protéine mature, puisque le leader est clivé au cours du mécanisme d'insertion dans la membrane.

La **pré-séquence** se distingue de la **pro-séquence** qui décrit les régions supplémentaires présentes sur les protéines qui existent sous forme de précurseurs stables. Quelques protéines peuvent posséder les deux. Par exemple, l'insuline est d'abord synthétisée sous forme de **préproinsuline**. La **pré-séquence** est clivée précocement au cours de la sécrétion, engendrant la **proinsuline** qui est le substrat de l'étape ultérieure de la maturation de l'insuline.

Le rôle de la séquence leader est conçu pour s'adapter aux circonstances. Pour certaines protéines synthétisées dans le cytoplasme mais destinées à résider dans le chloroplaste ou les mitochondries, le procédé est **post-traductionnel**. Le produit de synthèse protéique cytoplasmique est un précurseur plus long d'environ 45 acides aminés que la protéine mature. Le précurseur entier est libéré des polysomes. Quand il est ajouté à des organelles intactes, il peut être incorporé dans le compartiment.

Comme la **Figure 8.13** l'illustre, la séquence leader guide le précurseur à travers la membrane de l'organelle.

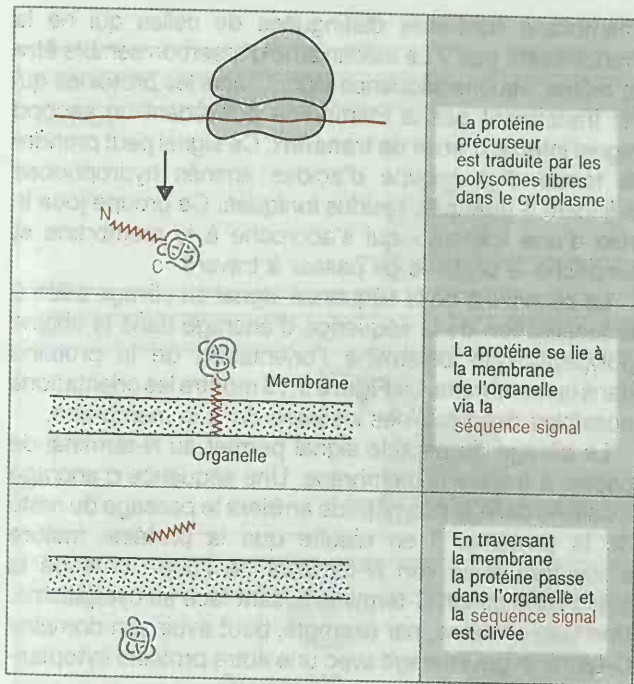


Figure 8.13. Les séquences signal permettent aux protéines de reconnaître les membranes des mitochondries ou des chloroplastes.

Au cours de ce passage, la séquence leader est clivée par une protéase associée à la membrane mitochondriale. Le rôle du leader est d'être reconnu par la membrane de l'organelle, déclenchant le passage et la réaction de clivage. Il faut deux passages pour les protéines qui résident dans le compartiment interne. Il est à noter qu'un leader sensible au clivage n'est pas le seul moyen de conférer un tel ciblage ; quelques protéines mitochondriales sont reconnues sous leur forme mature et peuvent avoir une séquence interne qui peut permettre le passage à travers la membrane sans clivage.

En ce qui concerne les protéines qui sont incorporées dans d'autres membranes cellulaires, l'association commence le plus souvent pendant la traduction. Les polysomes qui synthétisent ces protéines sont associés aux membranes du réticulum endoplasmique. Les préprotéines ne sont pas libérées dans le cytoplasme sous forme d'un pool de précurseurs, mais passent directement du ribosome à travers la membrane. A partir des membranes du réticulum endoplasmique, les protéines entrent dans l'appareil de Golgi et sont ensuite dirigées vers leur destination finale, telle que le lysosome, une vésicule sécrétoire ou la membrane plasmique (les immunoglobulines et un grand nombre d'hormones empruntent une

telle voie pour sortir de la cellule). Le mécanisme par lequel les protéines sont synthétisées en contact direct avec la membrane est appelé **transfert cotraductionnel**.

Un modèle du mécanisme de l'insertion dans les membranes s'est basé sur des travaux utilisant le système microsomal eucaryotique (système qui contient les ribosomes et les membranes du réticulum endoplasmique). Ces microsomes peuvent empaqueter les protéines naissantes à l'intérieur de membranes, mais ils ne sont pas fonctionnels quand les préprotéines sont ajoutées après la traduction.

Selon l'**hypothèse du signal**, le leader de la protéine sécrétée constitue une **séquence signal** dont la présence détermine l'insertion membranaire. A de rares exceptions près, le N-terminal d'une protéine sécrétée est constitué d'un leader clivable de 16 à 29 acides aminés, commençant par deux ou trois résidus polaires suivis d'une forte proportion d'acides aminés hydrophobes ; il n'y a pas d'autre conservation de séquence notable.

La séquence signal permet aux ribosomes en cours de traduction de l'ARNm de s'attacher à la membrane. La responsabilité de la liaison initiale repose uniquement sur la séquence signal ; le ribosome se lie simplement grâce à la synthèse de la protéine sécrétée. *Il n'y a donc aucune différence intrinsèque entre les ribosomes de la fraction libre et les ribosomes de la fraction liée à la membrane.*

La séquence signal est reconnue en tant que signal de liaison par le ribosome sans doute grâce à son caractère hydrophobe, dès que les premiers acides aminés du N-terminal sont synthétisés. Quand la chaîne protéique est bien insérée dans la membrane, la séquence signal peut être clivée par une protéase située dans la membrane. Tandis que le ribosome achève la traduction, la protéine est déjà bien engagée au travers de la membrane. Le point critique de ce mécanisme est qu'une protéine ne peut s'insérer dans la membrane cependant qu'elle est synthétisée.

Les membranes ayant subi des lavages salins ne permettent pas l'attachement du ribosome ; mais cette propriété peut être retrouvée en leur ajoutant ultérieurement le milieu salin. Le constituant actif est désigné sous le nom de **particule de reconnaissance du signal (SRP pour signal recognition particle en anglais)**. De structure cylindrique, 5-6 nm de largeur et 23-24 nm de long, elle est isolée comme un complexe ribonucléoprotéique 11S, qui contient 6 protéines (d'une masse totale de 240 000 daltons) et un petit ARN 7S (305 bases, 100 000 daltons). L'ARN 7S constitue le squelette de la particule ; son absence bloque l'agrégation des protéines individuelles.

La SRP a deux propriétés. Elle peut se lier, d'une part, à la séquence signal des protéines sécrétoires naissantes

et, d'autre part, à un récepteur protéique situé dans la membrane. On peut reconstituer son activité *in vitro* à partir de ses constituants isolés. En fait, une SRP fonctionnelle peut être obtenue avec un ARN 7S et des protéines n'ayant pas la même origine. Comme le reste de l'appareil qui transporte et traite les protéines membranaires, la SRP est bien conservée chez tous les eucaryotes.

La signal peptidase (identifiée *in vitro*) est constituée d'un complexe de six protéines. Une de ces protéines possède sans doute l'activité peptidase et les autres interviennent dans sa localisation dans la membrane. Ce complexe est plus abondant que la SRP et le récepteur SRP (qui agit de façon catalytique). Il équivaut approximativement à la quantité de ribosomes liés et peut donc avoir une fonction structurale.

Les étapes de la traduction des protéines membranaires sont résumées dans la **Figure 8.14**. Le rôle de la SRP est de transférer le ribosome vers un autre composant de la membrane. Dans ce mécanisme, le récepteur du SRP est un intermédiaire. C'est une protéine de 72 000 daltons, localisée dans le réticulum endoplasmique. Son rôle dans la translocation protéique semble être transitoire. Quand la SRP se lie à la séquence signal, la traduction s'arrête. Quand la SRP se lie à son récepteur, elle libère le ribosome qui est « ciblé » vers un autre constituant de la membrane (non encore identifié). Dans ce cas, la traduction peut reprendre. La SRP et son récepteur ont accompli leur rôle quand le ribosome a été transmis à cet autre constituant.

Comment une protéine passe-t-elle à travers la membrane ? L'idée première était que le N-terminal indiquait la voie à suivre comme l'illustre la **Figure 8.14**. L'autre possibilité cependant, était que le N-terminal entrait mais ne passait pas nécessairement au travers ; en fait la séquence signal pourrait s'enfoncer dans la phase lipidique de la membrane tandis que le reste du polypeptide en cours de synthèse se logerait dans le réticulum endoplasmique. En tout cas, la séquence signal ne peut exercer sa fonction qu'au cours de la traduction et non en tant que préprotéine mature.

Certaines protéines n'ont pas de séquence signal identifiable mais sont tout de même insérées dans les membranes. Quelques-unes, sans doute, possèdent une séquence hydrophobe qui favorise le passage cotraductionnel bien qu'elle ne soit pas clivée. D'autres peuvent être liées aux lipides de façon covalente ; la lipoprotéine qui en résulte possède une affinité inhérente pour les membranes.

Comment les protéines qui sont sécrétées à travers la

membrane sont-elles distinguées de celles qui ne la franchissent pas ? Le mécanisme d'insertion semble être le même, via une séquence signal, mais les protéines qui ne franchiront pas la membrane possèdent un second signal interne d'**arrêt de transfert**. Ce signal peut prendre la forme d'un groupe d'acides aminés hydrophobes adjacent à quelques résidus ioniques. Ce groupe joue le rôle d'une « ancre » qui s'accroche à la membrane et empêche la protéine de passer à travers.

La sensibilité de la séquence signal au clivage alliée à la localisation de la séquence d'ancrage dans la chaîne polypeptidique détermine l'orientation de la protéine dans la membrane. La **Figure 8.15** montre les orientations possibles des protéines insérées dans la membrane.

Le clivage du peptide signal permet au N-terminal de passer à travers la membrane. Une séquence d'ancrage localisée dans le polypeptide arrêtera le passage du reste de la protéine. Il en résulte que la protéine mature s'orientera avec son N-terminal de l'autre côté de la membrane et son C-terminal faisant face au cytoplasme. Une telle protéine, par exemple, peut avoir son domaine C-terminal qui interagit avec une autre protéine cytoplasmique.

L'orientation opposée est le résultat de situations où les fonctions du signal et de l'ancre sont combinées dans un seul groupe d'acides aminés (non clivable). De telles protéines pénètrent dans les membranes de façon conventionnelle mais aboutissent à une orientation inverse car la progression du N-terminal est bloquée par l'ancrage du signal d'arrêt tandis que le C-terminal passe à travers la membrane. Ainsi le C-terminal de la protéine mature sera de l'autre côté de la membrane tandis que le N-terminal restera dans le cytoplasme.

L'hypothèse du signal vaut aussi pour les bactéries. Les protéines qui seront exportées peuvent être munies de séquences leader N-terminales, avec un N-terminal hydrophile et un cœur adjacent hydrophobe. Elles peuvent être synthétisées par les ribosomes liés aux membranes. Les propriétés d'un système de transport *in vitro* des protéines à travers les membranes bactériennes laissent supposer que le cytoplasme faisant face à la membrane interne contient une ou des protéines nécessaires au mécanisme du transport et qu'elles pourraient reconnaître la séquence signal.

La sécrétion peut être inhibée par des mutations sur les leaders N-terminaux ; elles peuvent être supprimées par des mutations sur d'autres gènes qui sont ainsi définis comme des constituants de l'appareil servant à exporter des protéines. Comme chez les eucaryotes, des signaux secondaires peuvent être nécessaires à une localisation

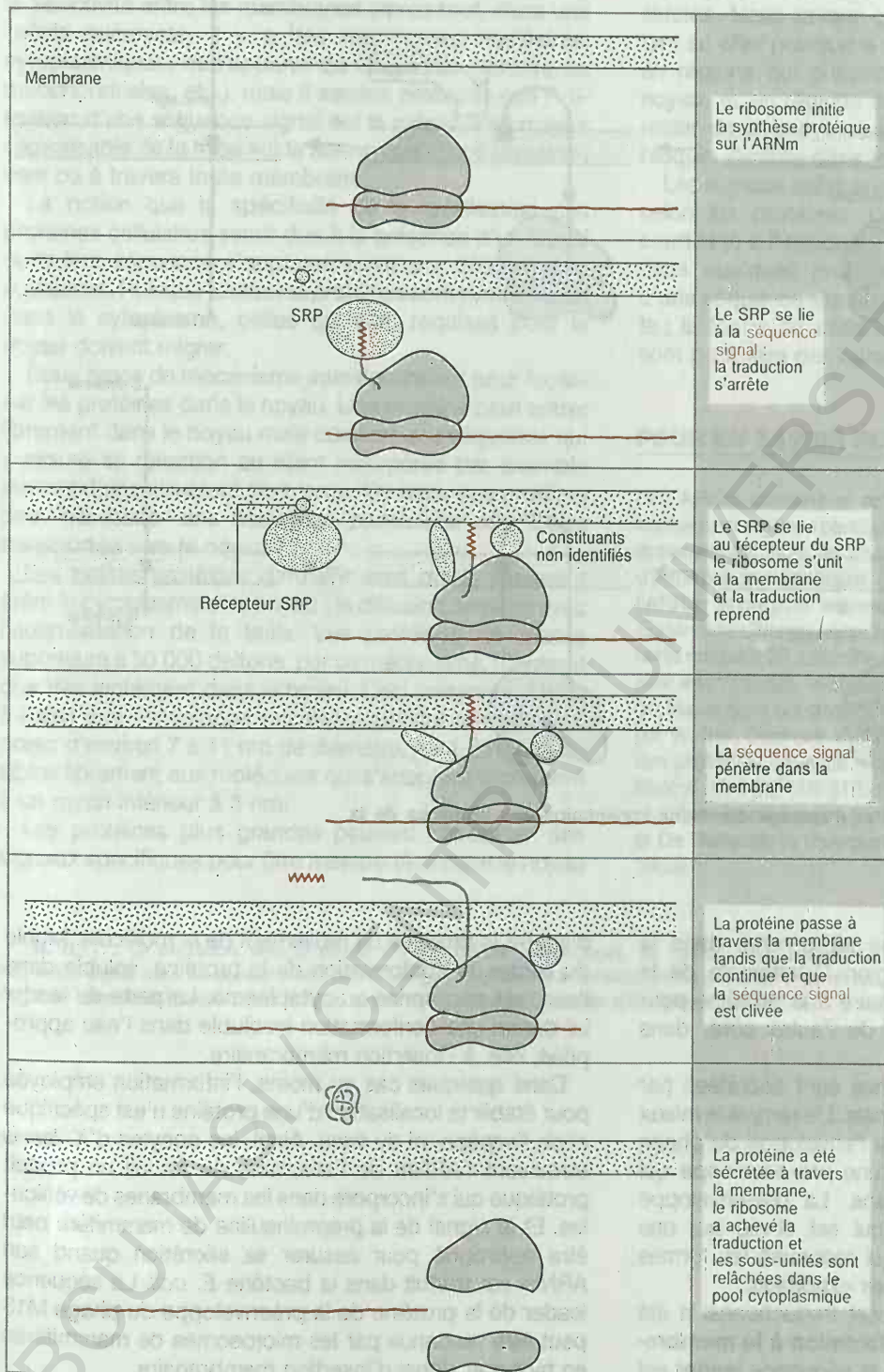


Figure 8.14
L'hypothèse du signal propose que les ribosomes qui synthétisent les protéines sécrétées sont unis aux membranes par la séquence signal du polypeptide naissant.

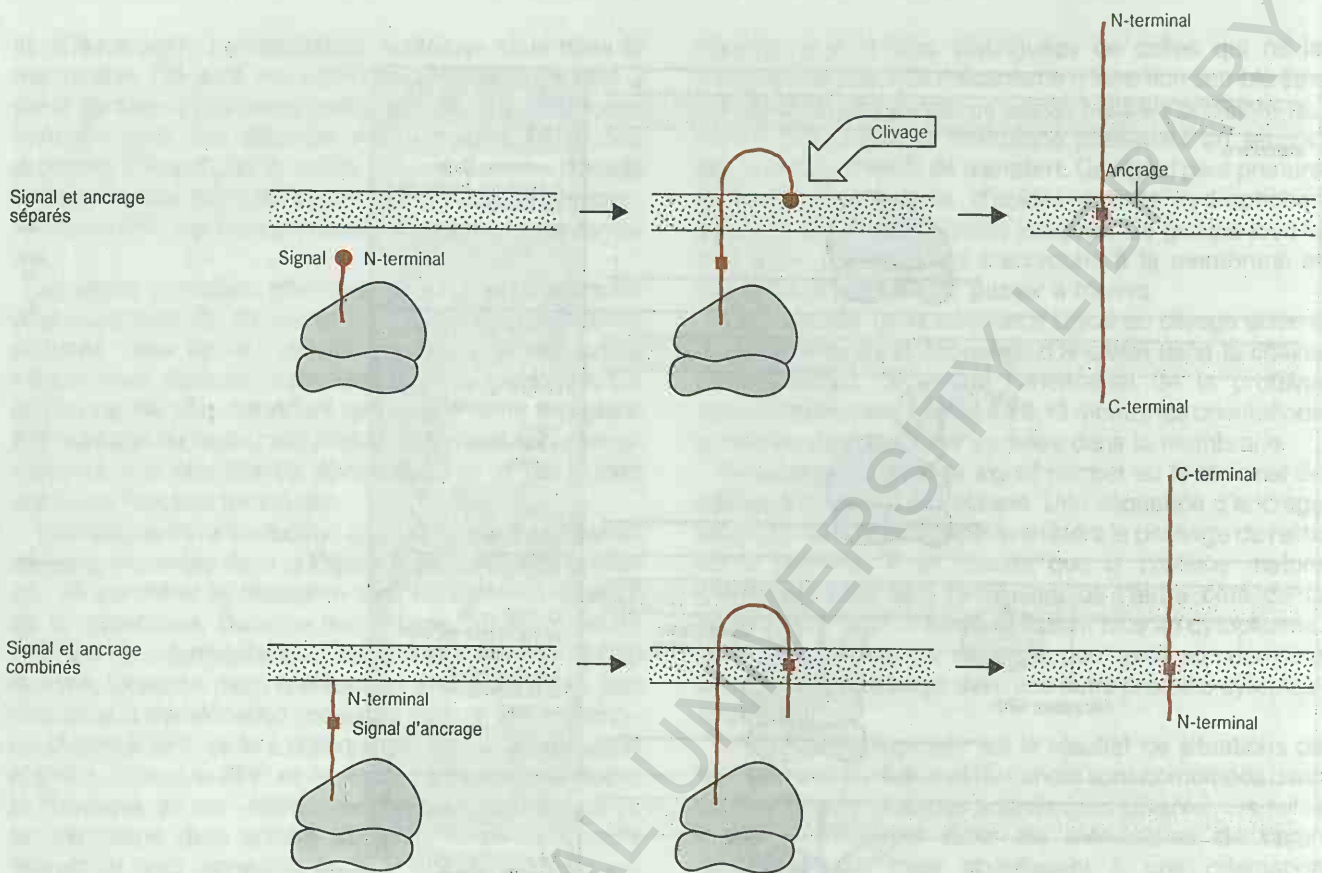


Figure 8.15

L'organisation des séquences signal et d'ancrage détermine l'orientation des protéines de la membrane.

adéquate après que la protéine ait pénétré dans la membrane. Par exemple, la région C-terminale de la β -lactamase d'*E. coli* est nécessaire à la protéine pour quitter la membrane et entrer, de l'autre côté, dans l'espace périplasmique.

Certaines protéines bactériennes sont sécrétées par des mécanismes post-traductionnels. L'exemple le mieux étudié est celui de la protéine de l'enveloppe du phage M13, synthétisée sous forme d'une pré-enveloppe qui peut s'insérer dans la membrane. La pré-enveloppe contient une séquence leader qui est clivée par une enzyme, la **leader peptidase**, qui reconnaît les formes précurseurs de plusieurs protéines exportées.

Un rôle éventuel du clivage post-traductionnel a été résumé dans l'hypothèse de l'**affectation à la membrane**. Selon ce modèle, le rôle de la séquence leader est

d'altérer le procédé de repliement de la molécule. Munie du leader, la conformation de la protéine, soluble dans l'eau, est appropriée au cytoplasme. La perte du leader lui donne une conformation insoluble dans l'eau appropriée, elle, à l'insertion membranaire.

Dans quelques cas au moins, l'information employée pour établir la localisation d'une protéine n'est spécifique ni de l'espèce, ni du tissu. Ainsi, les oocytes d'*X. laevis* traduisent l'ARNm de l'albumine de rat en un produit protéique qui s'incorpore dans les membranes de vésicules. Et le signal de la préproinsuline de mammifère peut être approprié pour assurer sa sécrétion quand son ARNm est traduit dans la bactérie *E. coli*. La séquence leader de la protéine de la préenveloppe du phage M13 peut être reconnue par les microsomes de mammifères en tant que signal d'insertion membranaire.

Il reste encore beaucoup à apprendre sur la base de la sélectivité entre les membranes (après tout, dans une cellule eucaryote, il y a des membranes nucléaires, cytoplasmiques, de l'appareil de Golgi, des membranes mitochondriales, etc.), mais il semble probable que l'utilisation d'une séquence signal est le mécanisme majeur responsable de la mise sur la bonne voie d'une protéine : vers ou à travers toute membrane.

La notion que la spécificité de la localisation des protéines cellulaires serait due à la présence d'un signal dans leur séquence s'applique aussi aux déterminants nucléaires. Puisque toutes les protéines sont synthétisées dans le cytoplasme, celles qui sont requises pour le noyau doivent migrer.

Deux types de mécanisme interviendraient pour localiser les protéines dans le noyau. Une protéine peut entrer librement dans le noyau mais contenir une séquence qui y assure sa rétention en étant incorporée par exemple au sein d'une structure plus large. Ou bien, une protéine peut nécessiter une séquence particulière afin d'être transportée vers le noyau*.

Les petites protéines diffusent sans doute librement entre le cytoplasme et le noyau ; la diffusion diminue avec l'augmentation de la taille. Les protéines de masse supérieure à 50 000 daltons, par ce mécanisme, n'entrent que très lentement dans le noyau. Ceci correspond bien à l'idée que l'enveloppe nucléaire contient une série de pores d'environ 7 à 11 nm de diamètre, peut-être accessibles librement aux molécules qui s'adaptent facilement à un rayon inférieur à 5 nm.

Les protéines plus grandes peuvent nécessiter des signaux spécifiques pour être transportées dans le noyau

bien que leur mécanisme d'entrée ne soit pas encore élucidé. Nous savons que des signaux spécifiques existent en effet puisque le clivage peut scinder les protéines en régions qui possèdent le signal et entrent dans le noyau, et en régions dépourvues de signal qui doivent rester dans le cytoplasme. La mutation du signal peut bloquer l'entrée dans le noyau de la protéine entière.

Les signaux utiles sont situés dans différentes régions, selon les protéines. Des séquences d'acides aminés courtes (5 à 7 résidus) plutôt basiques, ont été identifiées dans quelques protéines. Cependant, la conservation d'une séquence signal nucléaire n'est pas encore évidente ; la forme de cette région ou son caractère basique sont peut-être des caractéristiques importantes.

POUR EN SAVOIR PLUS

Les ARNm bactériens ont été revus principalement dans le contexte des gènes particuliers qu'ils représentent ; les citations apparaissent dans les chapitres 10 et 11. La principale source d'informations générales concernant la structure et le rôle de l'ARNm eucaryote est celle de **Lewin**, *Gene Expression*, 2, *Eucaryotic Chromosomes* (Wiley, New York, 1980), en particulier le chapitre 23. La coiffe a été revue par **Bannerjee** (*Microbiol. Rev.* 44, 175-205, 1980). Les mécanismes de la localisation des protéines sont controversés. L'hypothèse du signal a été revue par **Walter, Gilmore & Blobel** (*Cell* 38, 5-8, 1984). Le transport des protéines dans les mitochondries a été revu par **Schatz & Butow** (*Cell* 32, 316-318, 1983). **Oliver** a revu l'exportation des protéines bactériennes (*Ann. Rev. Microbiol.* 39, 615-648, 1985), et **De Robertis** le transport dans le noyau (*Cell* 32, 1021-1022, 1983).

* N. d. T. : La migration vers le noyau d'une protéine comporterait, au moins, deux étapes. La première, exigeant la présence d'une séquence-signal mais ne consommant pas d'énergie, aboutirait à la localisation de la protéine au niveau de la membrane nucléaire. La seconde étape — celle de la translocation dans le noyau — nécessiterait l'apport d'énergie, et donc d'ATP.

PARTIE 3

CONTRÔLE DE L'EXPRESSION DES GÈNES PAR LA TRANSCRIPTION

Si l'on s'en réfère strictement au concept structural, le génome est considéré comme une mosaïque de calques moléculaires indépendants pour l'élaboration de chaque constituant cellulaire. Lors de l'exécution de ces plans, la coordination est de toute évidence un impératif vital. La découverte de gènes régulateurs et opérateurs, et celle de la régulation négative de l'activité des gènes de structure révèlent que le génome possède non seulement une série de calques mais aussi un programme coordonné de synthèse protéique et le moyen de contrôler son exécution.

François Jacob & Jacques Monod, 1961

PARTIE 3

L'EXPRESSION DES GÈNES
PAR LA TRANSCRIPTION
ET LE CONTRÔLE DE LA TRANSCRIPTION

Éditions de la Faculté de Médecine, 1987

CHAPITRE 9

LES INTERACTIONS ARN POLYMÉRASE-PROMOTEUR CONTRÔLENT L'INITIATION

Comment l'ARN devint-il l'intermédiaire entre l'ADN et la protéine ? Il est possible que les cellules primitives ne faisaient pas de distinction entre les différents types d'acides nucléiques ; donc, ce qui était considéré comme le génome était impliqué, directement, à la fois dans la réplication et dans la traduction. A un certain moment, il a pu devenir avantageux de séparer la traduction du génome, de façon à ce que les protéines soient synthétisées sur des messagers distincts du matériel génétique lui-même.

Il est impossible de dire comment cette séparation fut liée chronologiquement au développement des autres acides ribonucléiques impliqués dans la traduction, mais il est frappant de constater que l'ARN est présent dans les ribosomes de même qu'il constitue l'adaptateur ARNt. (Il ne serait pas surprenant que le rôle de l'ARN dans les ribosomes fût plus important dans le passé qu'il n'apparaît aujourd'hui.) Peut-être l'ARN était-il l'acide nucléique originel, et son ubiquité actuelle est-elle simplement une réminiscence de son activité passée.

Tout ceci pour dire que l'ARN joue un rôle central dans l'expression des gènes, non seulement en constituant le messager, mais également en fournissant le moyen de sa traduction en protéine. En un sens, ces rôles représentent les différents intérêts spécifiques d'un conglomerat

d'ARN impliqués dans l'expression des gènes en général, mais possédant plusieurs fonctions différentes.

La production de chaque type d'ARN a une origine commune : la transcription de l'ADN. Dans le cas de l'ARNm, le produit est un intermédiaire dont la fonction requiert une traduction. Dans le cas de l'ARNt et de l'ARNr, le produit de transcription remplit lui-même la fonction finale.

Transcrire ou ne pas transcrire ? c'est la question. La transcription est la principale étape pendant laquelle l'expression des gènes est contrôlée. La première (et parfois la seule) étape du contrôle est la décision de transcrire ou non le gène. En considérant les différentes étapes de la transcription, nous devrons dorénavant garder à l'esprit les opportunités qu'elles offrent pour réguler l'activité des gènes.

(La transcription n'est pas le seul moyen par lequel l'ARN peut être synthétisé. Des virus possédant des génomes à ARN produisent des enzymes capables de synthétiser de l'ARN sur une matrice elle-même constituée d'ARN. De telles réactions produisent des ARNm codant pour des protéines nécessaires au cycle infectieux [transcription d'ARN] et fournissent de l'ARN génomique pour perpétuer ce cycle [réplication d'ARN]. Les rétrovirus présentent une réaction supplémentaire dans laquelle

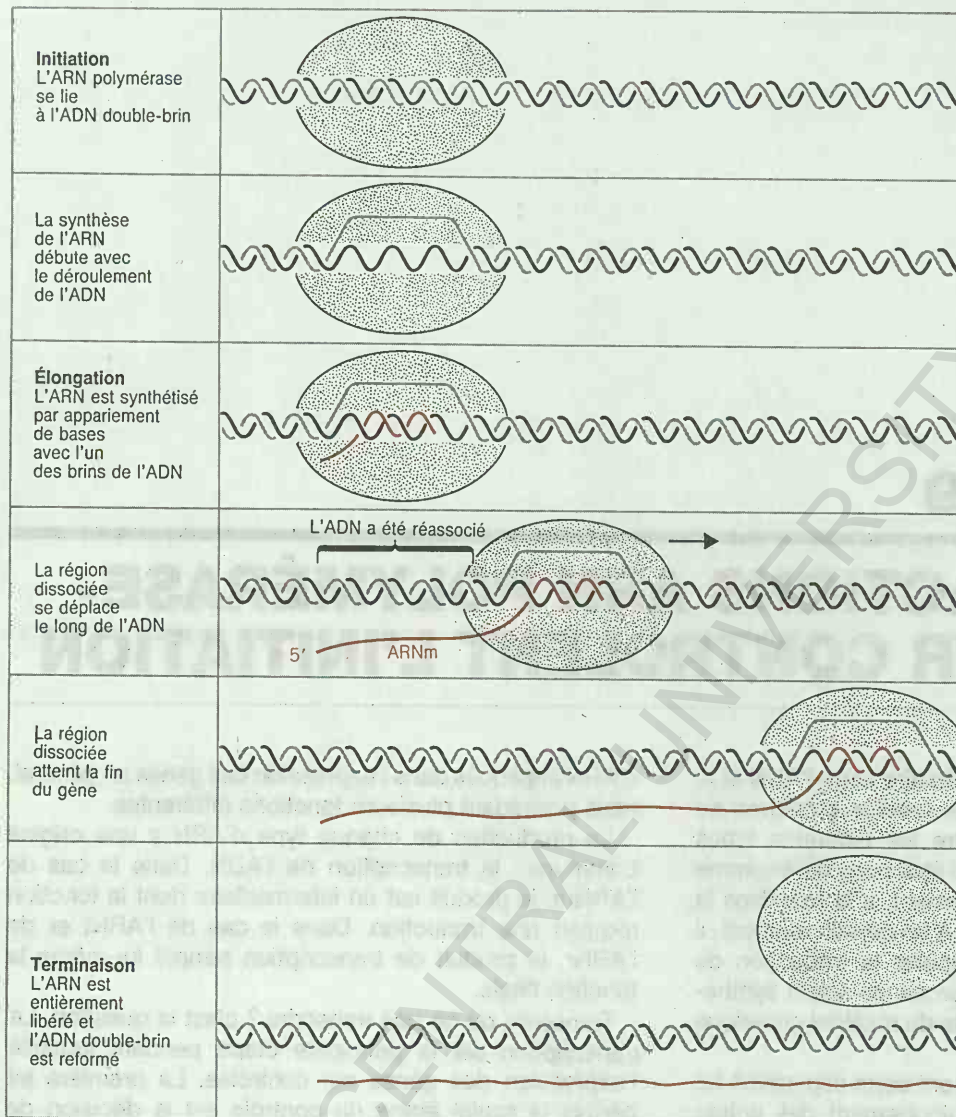


Figure 9.1

L'ARN est synthétisé par appariement de base avec l'un des brins de l'ADN dans une région dissociée de façon transitoire. Lorsque la région dissociée se déplace, la double hélice d'ADN se reforme derrière elle, déplaçant l'ARN sous forme d'une simple chaîne polynucléotidique.

l'ARN viral sert de matrice pour la transcription inverse afin de produire un ADN complémentaire.)

LA TRANSCRIPTION EST CATALYSÉE PAR L'ARN POLYMÉRASE

La transcription implique la synthèse d'une chaîne d'ARN qui représente l'un des deux brins d'ADN. Par « représen-

ter », nous voulons dire que l'ARN est identique par sa séquence à l'un des brins d'ADN ; il est complémentaire de l'autre brin qui lui sert de matrice pour sa synthèse.

La transcription se déroule par le procédé habituel d'appariement de bases complémentaires, catalysée par l'enzyme **ARN polymérase**. La réaction peut être divisée en trois étapes illustrées par la Figure 9.1.

- **L'initiation** débute avec la fixation de l'ARN polymérase à l'ADN double brin. Pour que l'appariement des

bases entre le brin matriciel et les ribonucléotides soit possible, les brins d'ADN doivent être séparés. La dissociation est un événement local qui débute au site de fixation de l'ARN polymérase. L'étape d'initiation est achevée lorsque le premier nucléotide est incorporé.

La séquence complète d'ADN nécessaire à ces réactions est appelé le **promoteur**. Le site sur lequel le premier nucléotide est incorporé est appelé le **site d'initiation** ou le **point d'initiation**.

- **L'élongation** décrit la phase pendant laquelle les nucléotides sont ajoutés de façon covalente à l'extrémité 3' de la chaîne polynucléotidique en croissance. Des bases successives viennent s'ajouter à la chaîne d'ARN qui forme un hybride ARN-ADN dans la région dissociée.

Pour poursuivre la synthèse, l'enzyme va se déplacer le long de l'ADN en dissociant la double hélice pour exposer un nouveau segment de la matrice à l'état de simple brin. En se déplaçant, l'ARN précédemment synthétisé est écarté du brin matriciel de l'ADN, qui se réassocie avec son partenaire d'origine pour reformer la double hélice.

Ainsi l'élongation implique le déplacement le long de l'ADN d'un court segment transitoirement dissocié, existant sous forme d'un hybride ARN-ADN et d'un ADN simple brin déplacé.

- **La terminaison** implique la reconnaissance du point à partir duquel plus aucune base ne sera ajoutée à la chaîne. Pour terminer la transcription, la formation de liaisons phosphodiester doit cesser et le complexe transcriptionnel doit se dissocier. Lorsque la dernière base est ajoutée à la chaîne d'ARN, l'hybride ARN-ADN est rompu, l'ADN retrouve son état de double brin et, l'enzyme et l'ARN en sont tout deux libérés. La séquence d'ADN nécessaire à ces réactions est appelée le **terminateur**.

Initialement définie simplement par sa capacité à incorporer des nucléotides à l'ARN sous la direction de la matrice d'ADN, l'enzyme ARN polymérase est maintenant considérée comme faisant partie d'un appareil plus complexe impliqué dans la transcription. La capacité de catalyser la synthèse d'ARN définit le composant minimum pouvant être décrit comme ARN polymérase. Il contrôle l'appariement des bases du substrat ribonucléotidique avec l'ADN et catalyse la formation de liaisons phosphodiester entre elles.

Des activités auxiliaires peuvent être nécessaires à l'initiation et à la terminaison de la synthèse d'ARN, lorsque l'enzyme doit être associée ou dissociée d'un

site spécifique de l'ADN. L'analogie avec la répartition des tâches entre le ribosome et les facteurs de la synthèse protéique est manifeste. Il est parfois difficile de dire si une protéine particulière, impliquée dans l'une des étapes de la transcription, doit être considérée comme un élément de « l'ARN polymérase » ou comme un facteur auxiliaire.

Tous les composants impliqués dans l'élongation sont nécessaires à l'initiation et à la terminaison. Les gènes peuvent différer dans leur dépendance vis-à-vis de polypeptides additionnels aux étapes d'initiation ou de terminaison. Certains de ces polypeptides peuvent être nécessaires à tous les gènes mais d'autres peuvent n'être spécifiquement indispensables qu'à l'initiation ou à la terminaison de gènes particuliers. Un polypeptide additionnel, nécessaire à la reconnaissance de tous les promoteurs (ou de tous les sites de terminaisons) est plutôt à classer comme faisant partie de l'enzyme. Un polypeptide nécessaire uniquement à l'initiation (ou à la terminaison) de gènes particuliers est plutôt à classer comme facteur auxiliaire de contrôle.

Avec les enzymes bactériennes, il est possible de commencer à définir les rôles de chaque polypeptide dans les différentes étapes de la transcription. Chez les eucaryotes, les enzymes sont moins bien purifiées et les activités enzymatiques actuelles ont encore à être complètement purifiées à partir de préparations relativement grossières. Assez paradoxalement, chez les eucaryotes nous avons commencé à isoler des facteurs auxiliaires nécessaires à initier ou à terminer des gènes particuliers, alors que la préparation de base de la polymérase elle-même est encore plutôt mal caractérisée.

L'ARN POLYMERASE BACTÉRIENNE CONSISTE EN UNE ENZYME « CŒUR » ET LE FACTEUR SIGMA

Un seul type d'ARN polymérase est responsable de toutes les synthèses d'ARNm, ARNr et ARNt dans la bactérie. Le nombre total de molécules d'ARN polymérase présentes dans une cellule d'*E. coli* est d'environ 7 000. Beaucoup d'entre elles transcrivent continuellement. 2 000 à 5 000 enzymes probablement synthétisent l'ARN à chaque instant, leur nombre dépendant des conditions de croissance.

L'ARN polymérase la mieux caractérisée est celle d'*E. coli*, mais sa structure est similaire chez toutes les autres bactéries étudiées. L'enzyme complète ou **holoenzyme** a une masse moléculaire d'environ 480 000 daltons. Sa composition en sous-unités est résumée dans le **Tableau 9.1**.

Tableau 9.1

L'ARN polymérase d'*E. coli* a quatre types de sous-unités.

Sous-unité	Gène	Nombre	Masse (daltons)	Localisation	Fonctions possibles
α	<i>rpoA</i>	2	40 000 chacune	enzyme « cœur »	liaison du promoteur
β	<i>rpoB</i>	1	155 000	enzyme « cœur »	liaison des nucléotides
β'	<i>rpoC</i>	1	160 000	enzyme « cœur »	liaison de la matrice
σ	<i>rpoD</i>	1	85 000	facteur sigma	initiation

Les sous-unités α , β et β' ont une taille plutôt constante chez les différentes espèces bactériennes ; la sous-unité σ varie plus largement de 32 000 à 92 000. L'enzyme a une structure plutôt allongée, avec une dimension maximale de 15 nm (un tour d'hélice d'ADN représente 3,4 nm).

L'holoenzyme ($\alpha_2\beta\beta'\sigma$) peut être séparée en deux composants, l'**enzyme « cœur »** (ou enzyme minimum) ($\alpha_2\beta\beta'$) et le **facteur sigma** (le polypeptide σ). Les noms donnés reflètent le fait que seule l'holoenzyme peut initier la transcription ; et qu'ensuite le « facteur » sigma est libéré, laissant l'enzyme « cœur » prendre en charge l'élongation. Ainsi l'enzyme « cœur » a la possibilité de synthétiser l'ARN sur une matrice d'ADN, mais ne peut initier la transcription aux sites corrects.

L'enzyme « cœur » débute la transcription à l'endroit où les brins de l'ADN d'un complexe d'initiation sont dissociés. Lorsque l'enzyme se déplace le long de la matrice, allongeant la chaîne d'ARN, la région de dissociation locale se déplace avec elle. L'enzyme couvre environ 60 pb d'ADN ; le segment dissocié n'occupe qu'une assez petite partie de cette longueur, inférieure à 17 pb en fonction du degré global de déroulement.

Lorsque l'ADN se déroule pour libérer la matrice, chacun de ses brins entre en contact probablement avec un site distinct de la structure de l'enzyme. Comme l'indique la **Figure 9.2**, le brin matriciel sera libéré un peu en amont du point où le ribonucléotide sera ajouté à la chaîne d'ARN, il se présentera sous forme d'hybride ADN-ARN dans la région où l'ARN a été nouvellement synthétisé. La longueur de la zone hybride est peut-être légèrement plus courte que la région d'ADN dissocié. Il est probable que l'hybride ARN-ADN ait une longueur d'environ 12 pb.

Lorsque l'enzyme quitte cette région, l'ADN double brin se reforme et l'ARN est libéré sous forme d'une chaîne de polynucléotides. Près de 50 ribonucléotides nouvellement ajoutés à la chaîne en croissance sont complexés à l'ADN et/ou à l'enzyme à chaque instant.

Nous ne comprenons toujours pas très bien la topologie de la dissociation et de la réassociation pendant la transcription, mais la faculté qu'ont les ARN polymérases purifiées de transcrire l'ADN double brin *in vitro* implique

que la réaction dépend d'une propriété intrinsèque de l'enzyme. La dissociation et la réassociation impliquent que les brins de l'ADN tournent l'un autour de l'autre. L'ADN tournant dans le sens du déroulement en avant de l'enzyme et tournant dans le sens inverse derrière elle, constituerait un mécanisme possible. Ceci pourrait requérir l'assistance *in vivo* d'autres activités enzymatiques pour ajuster la topologie de l'ADN.

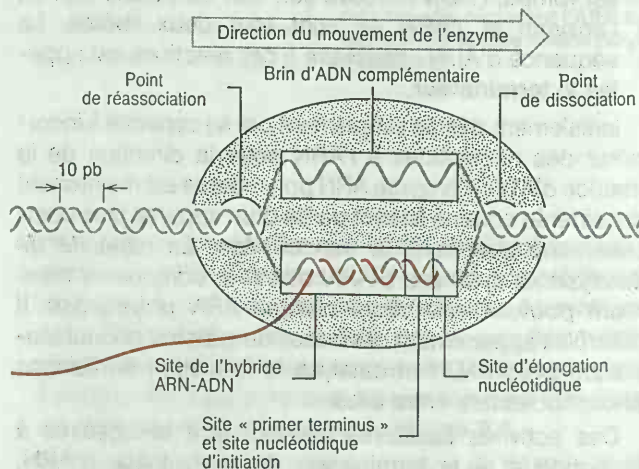
Figure 9.2

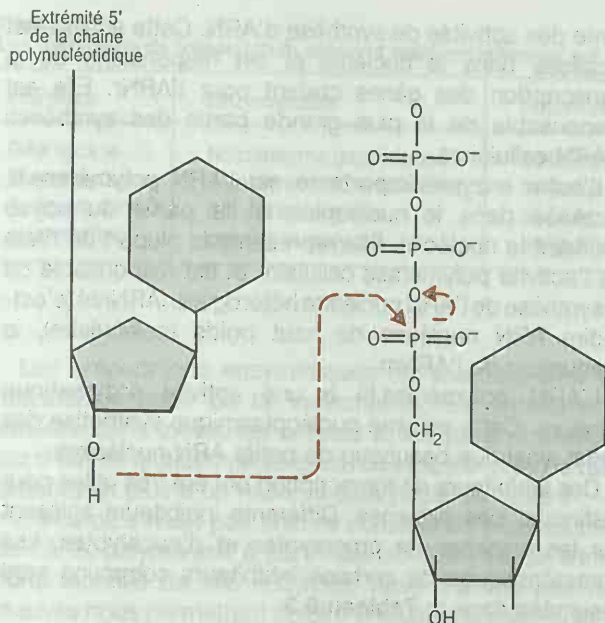
L'ARN polymérase bactérienne recouvre environ 60 pb d'ADN et possède plusieurs centres actifs.

Pendant l'élongation, les groupes réactionnels sont localisés en deux sites. L'endroit occupé par le nucléoside triphosphate entrant est le **site nucléotidique d'élongation**. La position du dernier nucléotide ajouté à la chaîne définit le site « **primer terminus** ».

Quand la transcription est initiée, il n'y a pas bien sûr de « primer terminus » et le tout premier nucléotide (généralement une purine) entre en contact avec le **site nucléotidique d'initiation**, qui doit largement chevaucher le site « primer terminus ». Le premier nucléotide incorporé à la chaîne garde tous ses résidus 5' triphosphate.

La longueur de la zone dissociée est exagérée dans l'illustration (environ 4 fois).



**Figure 9.3**

La formation d'une liaison phosphodiester implique l'attaque hydrophile par le groupe 3'-OH du dernier nucléotide de la chaîne sur le groupe 5' triphosphate du nouveau nucléotide, avec relargage de pyrophosphate.

Tous les acides nucléiques sont synthétisés à partir de nucléosides précurseurs 5' triphosphates. La **Figure 9.3** montre la réaction de condensation entre le groupe 3'-OH du dernier nucléotide incorporé dans la chaîne et le groupe 5' triphosphate du nouveau nucléotide. Celui-ci perd ses deux phosphates terminaux (γ et β) ; son groupe α est utilisé dans la liaison phosphodiester, le liant au nucléotide précédent.

L'enzyme « cœur » doit maintenir les deux groupes réactifs dans la position propre à la formation de la liaison phosphodiester. Quand ils sont liés de façon covalente, l'enzyme « cœur » se déplace d'une base le long de la matrice d'ADN, ainsi la réaction peut à nouveau avoir lieu. La vitesse de réaction est rapide, environ 40 nucléotides par seconde à 37 °C (voir Chapitre 8).

Un nouveau nucléotide est accepté grâce à l'appariement de sa base avec le brin d'ADN matriciel, une action apparemment supervisée par l'enzyme. Le site a probablement une structure qui permet à la liaison phosphodiester d'avoir lieu uniquement lorsque le nucléotide s'apparie correctement avec la base de l'ADN. On peut

présumer que le nucléotide est rejeté s'il ne peut s'apparier ; un autre nucléotide peut alors prendre sa place.

Nos connaissances sur la topologie de l'enzyme « cœur » sont très primitives, et actuellement, nous ne pouvons faire au mieux qu'une représentation schématique des sites définis par les différentes fonctions enzymatiques, comme illustré dans la **Figure 9.2**. Aucun de ces sites n'a encore été physiquement localisé sur les sous-unités polypeptidiques. Quoiqu'il en soit, il existe quelques informations générales sur les rôles de chaque sous-unité.

Deux types d'antibiotiques agissent sur la sous-unité β , comme le montre la localisation de mutations conférant une résistance (voir Tableau 9.3). Les **rifamycines** (dont la rifampicine est la forme la plus utilisée) empêchent l'initiation, en agissant avant la formation de la première liaison phosphodiester. Les **streptolydignes** inhibent l'élongation de la chaîne. La sous-unité β est la cible de ces deux types d'antibiotiques. Elle est également marquée par certains analogues d'affinités des nucléosides triphosphates. Tous ces résultats suggèrent que la sous-unité β pourrait être impliquée dans la fixation du substrat nucléotidique.

L'héparine est un polyanion qui se lie à la sous-unité β' et inhibe la transcription *in vitro*. L'héparine entre en compétition avec l'ADN pour lier initialement la polymérase. La sous-unité β' est la plus basique, ce qui s'accorderait avec un rôle de ligand de la matrice.

La sous-unité α n'a pas de rôle connu. Quand le phage T4 infecte *E. coli*, la sous-unité α est modifiée par l'ADP-ribosylation d'une arginine. Cette modification est associée à une réduction d'affinité pour les promoteurs initialement reconnus par l'holoenzyme ; la sous-unité α pourrait ainsi jouer un rôle dans la reconnaissance du promoteur.

Ces attributions de fonctions individuelles sont très primitives. Il est probable que chaque sous-unité contribue à l'activité de l'enzyme « cœur » dans son ensemble et qu'il n'est pas possible de compartimenter ces actions.

Pourquoi l'ARN polymérase bactérienne requiert-elle une structure multimérique de si grande taille ? L'existence d'ARN polymérases beaucoup plus petites, ne comportant qu'une unique chaîne polypeptidique, codée par certains phages, démontre que le matériel nécessaire à la synthèse de l'ARN peut être beaucoup plus petit que celui constitué par l'enzyme de l'hôte.

Ces enzymes donnent une idée du matériel « minimum » nécessaire à la transcription. Ils ne reconnaissent que très peu de promoteurs sur l'ADN du phage et ils n'ont pas la possibilité de modifier le jeu de promoteurs

auquels ils répondent. Ainsi ils sont limités dans leur capacité intrinsèque à reconnaître très peu de séquences de liaisons spécifiques sur l'ADN et à synthétiser l'ARN. Quelle est leur complexité ?

Les ARN polymérases codées par les phages apparentés T3 et T7 sont des chaînes polypeptidiques simples d'environ 11 000 daltons chacune. Elles synthétisent l'ARN très rapidement (à un taux d'environ 200 nucléotides/seconde à 37 °C). La réaction d'initiation ne présente que très peu de variations.

Par contre, l'enzyme de la bactérie hôte peut transcrire n'importe laquelle des nombreuses unités de transcriptions (supérieures à 1 000). Certaines de ces unités sont transcrites directement, sans aucune aide additionnelle. Mais beaucoup d'unités ne peuvent être transcrites qu'en présence de facteurs protéiques supplémentaires. Certains de ces facteurs sont spécifiques d'une seule unité de transcription ; d'autres sont impliqués dans la transcription coordonnée de nombreuses unités. Certains phages induisent des changements généraux dans l'affinité de l'ARN polymérase de l'hôte ; celle-ci cesse alors de reconnaître les gènes de l'hôte et initie la transcription au niveau des promoteurs du phage.

Ainsi l'enzyme de l'hôte requiert la faculté d'interagir avec de nombreuses fonctions de l'hôte et du phage qui modifient ses activités transcriptionnelles intrinsèques. La complexité de l'enzyme, plutôt que les exigences inhérentes à son activité catalytique, peut donc, au moins en partie, refléter son besoin d'interagir avec de nombreux autres facteurs.

L'ARN POLYMERASE DES EUKARYOTES EST COMPOSÉE DE NOMBREUSES SOUS-UNITÉS

Le matériel de transcription des cellules eucaryotes est plus complexe et moins bien défini que celui des bactéries. Il y a trois ARN polymérases nucléaires, localisées dans différentes régions, chacune avec une structure complexe à sous-unités. *Chaque enzyme est responsable de la transcription de différentes classes de gènes.* Leurs propriétés générales sont définies dans le **Tableau 9.2**.

L'enzyme ARN polymérase I présente la plus impor-

tante des activités de synthèse d'ARN. Cette activité est localisée dans le nucléole et est responsable de la transcription des gènes codant pour l'ARNr. Elle est responsable de la plus grande partie des synthèses d'ARN cellulaires.

L'autre enzyme importante est l'ARN polymérase II, localisée dans le nucléoplasme (la partie du noyau excluant le nucléole). Elle représente la plupart du reste de l'activité polymérase cellulaire et est responsable de la synthèse de l'ARN nucléaire hétérogène (ARNnh) (c'est-à-dire ARN nucléaire de haut poids moléculaire), le précurseur de l'ARNm.

L'ARN polymérase III a une activité enzymatique mineure. Cette enzyme nucléoplasmique synthétise des ARNt ainsi que beaucoup de petits ARN nucléaires.

Des inhibiteurs de transcription ont été très utiles pour distinguer ces enzymes. Différents inhibiteurs agissent sur les enzymes de procaryotes et d'eucaryotes. Les caractéristiques de certains inhibiteurs communs sont résumées dans le **Tableau 9.3**.

Une distinction majeure entre les enzymes d'eucaryotes est faite à partir de leur réponse à l' α -amanitine, un octapeptide à double cycle. Dans des cellules d'origines aussi diverses que des cellules animales, végétales ou d'insectes, l'activité de l'ARN polymérase II est rapidement inhibée par de faibles concentrations d' α -amanitine. Quelle que soit l'origine des cellules, l'enzyme ARN polymérase I n'est pas inhibée. La réponse de l'ARN polymérase III à l' α -amanitine n'est pas aussi bien conservée ; dans les cellules animales, elle est inhibée par de fortes concentrations mais elle ne l'est pas chez la levure et chez les insectes.

Les activités enzymatiques brutes correspondent toutes à de grandes protéines, apparaissant comme des agrégats de 500 000 daltons ou plus. Leur composition en sous-unités est complexe. Chaque enzyme possède deux grandes sous-unités, généralement de 200 000 daltons environ pour l'une et de 140 000 daltons environ pour l'autre. Il existe moins de 10 sous-unités plus petites, dont la taille est comprise entre 10 000 et 90 000 daltons. Nous ignorons si certaines de ces sous-unités, trouvées dans les différentes enzymes, sont identiques ou non.

Tableau 9.2
Les noyaux des eucaryotes possèdent trois ARN polymérases.

Enzyme	Localisation	Production	Activité relative	Sensibilité à l' α -amanitine
ARN polymérase I	nucléole	ARN ribosomal	50-70 %	insensible
ARN polymérase II	nucléoplasme	ARNhn	20-40 %	sensible
ARN polymérase III	nucléoplasme	petits ARN	~ 10 %	selon les espèces

Tableau 9.3

Les inhibiteurs de transcription agissent préférentiellement sur des enzymes déterminées.

Inhibiteur	Enzyme cible	Action inhibitrice
Rifamycine Streptolydigne Actinomycine D α -amanitine	holoenzyme bactérienne « enzyme cœur » bactérien pol. I d'eucaryote pol. II d'eucaryote	se lie à β pour empêcher l'initiation se lie à β pour empêcher l'élongation se lie à l'ADN et empêche l'élongation se lie à l'ARN polymérase II

Les préparations enzymatiques représentent-elles le matériel de base de la transcription, essentiellement similaire dans toutes les cellules et sujette à régulation par d'autres facteurs protéiques, ou incluent-elles de tels facteurs en plus d'un matériel catalytique de base ?

Parce qu'il n'est pas encore possible de reconstituer de l'ARN polymérase à partir des sous-unités de n'importe laquelle de ces enzymes, nous n'avons pas de preuves nous permettant de dire si toutes les sous-unités protéiques font partie intégrante de chaque enzyme. Nous ne savons pas quelles sous-unités peuvent représenter des activités catalytiques et quelles autres peuvent être impliquées dans des fonctions de régulation.

Le moyen d'étudier cette question est d'utiliser des préparations d'enzymes isolées pour transcrire *in vitro* des matrices définies. On a caractérisé plusieurs réactions hétérologues, dans lesquelles une préparation d'ARN polymérase II d'un seul type cellulaire et d'une seule espèce est utilisée pour transcrire un gène qui est actif dans un autre type cellulaire d'une autre espèce. Le résultat positif de telles expériences indique que ni le tissu, ni des caractéristiques spécifiques aux espèces ne sont impliqués dans la reconnaissance du promoteur *per se*.

Bien sûr, cette conclusion n'exclut pas la possibilité que d'autres facteurs protéiques ou d'autres séquences ne soient impliqués dans la modulation de la réaction (spécialement pour accroître son efficacité) dans les conditions naturelles. On a trouvé des facteurs nécessaires à des ARN polymérases particulières pour initier la transcription d'un sous-ensemble de leurs gènes cibles. Des ensembles de facteurs peuvent être nécessaires pour transcrire des groupes particuliers de gènes.

La totalité de la complexité du matériel de transcription des eucaryotes est définie uniquement système par système ; tout ce que nous pouvons dire à présent est que des enzymes multimériques sont nécessaires pour chacune des trois classes de synthèse d'ARN. Il reste à voir si tous les composants de ces enzymes sont essentiels, et combien d'autres protéines sont nécessaires.

A cause de ces incertitudes, des systèmes relativement bruts et « sales » peuvent offrir plus de chances de caractériser la transcription *in vitro* que des systèmes purifiés et « propres » ; une purification trop poussée peut faire perdre spécifiquement ces facteurs que nous avons besoin de caractériser !

Les activités ARN polymérase de la mitochondrie et des chloroplastes semblent être de taille moindre et être distinctes des enzymes nucléaires. Le génome de ces organelles étant beaucoup plus petit, il est évident que leur polymérase n'a besoin de transcrire que quelques gènes, et que le contrôle de transcription devrait être beaucoup plus simple (si jamais il existe). Ainsi ces enzymes devraient être proches de celles du phage qui n'ont qu'un but défini et n'ont pas besoin de pouvoir répondre à un environnement plus complexe.

LE FACTEUR SIGMA DES BACTÉRIES CONTRÔLE LA LIAISON À L'ADN

La fonction du facteur sigma est d'assurer la liaison stable de l'ARN polymérase à l'ADN, *uniquement sur le promoteur et pas sur d'autres sites*.

L'enzyme « cœur » lui-même a de l'affinité pour l'ADN, affinité dans laquelle l'attraction électrostatique entre les protéines basiques et les acides nucléiques joue un rôle majeur. Il est probable que cette aptitude particulière à se lier à n'importe quel ADN, indépendamment de sa séquence propre, est une caractéristique de toutes les protéines possédant des sites spécifiques de liaison sur l'ADN (voir Chapitre 10).

Toute séquence d'ADN sur laquelle se fixe l'ARN polymérase dans cette réaction de liaison aspécifique est décrite comme un **site de liaison lâche**. Le complexe enzyme-ADN est décrit comme **fermé**, car l'ADN reste strictement sous sa forme double brin. Un complexe fermé est stable ; la demi-vie de dissociation du complexe enzyme-ADN est d'environ 60 minutes.

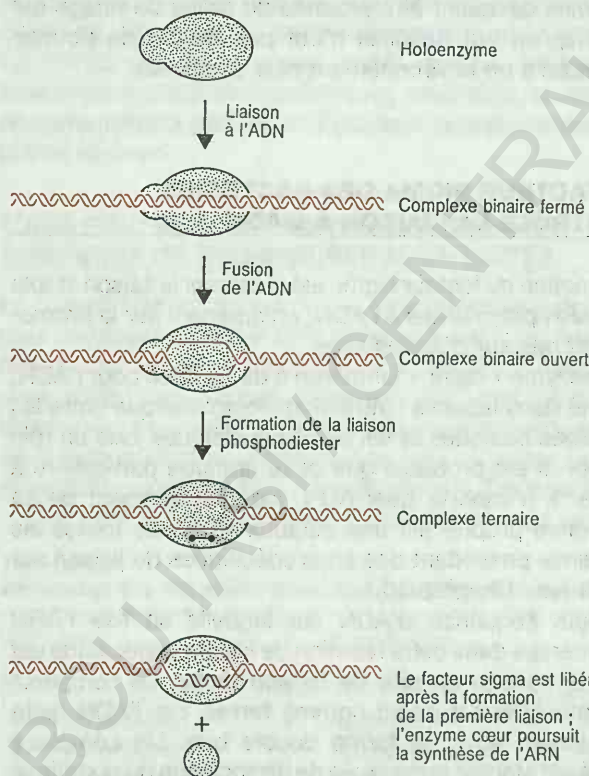
Le facteur sigma introduit un changement majeur dans l'affinité de l'ARN polymérase pour l'ADN. *L'holoenzyme a une capacité considérablement réduite de reconnaître les sites de liaisons lâches* — c'est-à-dire de se lier à n'importe quelle séquence d'ADN. La constante d'association de la réaction est diminuée d'un facteur de 10 000 environ et la demi-vie du complexe est inférieure à 1 seconde. Ainsi le facteur sigma déstabilise-t-il considérablement la capacité générale de liaison aspécifique.

Mais le facteur sigma confère aussi la possibilité de reconnaître des sites de liaisons spécifiques. L'holoenzyme se fixe très fortement au promoteur, avec une constante d'association augmentée par rapport à celle de l'enzyme « cœur » (en moyenne) de 1 000 fois et une demi-vie augmentée de plusieurs heures.

La constante d'association ne peut être considérée que comme une moyenne, car il y a (grossièrement) une variation d'un facteur cent dans la vitesse à laquelle

Figure 9.4

L'initiation de la transcription nécessite plusieurs étapes, durant lesquelles un complexe binaire fermé est converti en une forme ouverte puis en complexe ternaire.



l'holoenzyme se lie à différentes séquences de promoteur ; c'est un facteur important pour déterminer l'efficacité d'un promoteur à initier la transcription.

La reconnaissance de promoteurs par l'holoenzyme passe par deux étapes, illustrées dans la Figure 9.4. La réaction holoenzyme-promoteur débute de la même manière que la réaction de liaison lâche, en formant un complexe fermé. Mais celui-ci est alors converti en **complexe ouvert** par « fusion » d'une courte région d'ADN à l'intérieur de la séquence liée à l'enzyme. La série d'événements conduisant à la formation d'un complexe ouvert est appelée **liaison serrée**.

Les deux associations, fermée et ouverte, de l'ARN polymérase avec l'ADN sont décrites comme **complexes binaires**. L'étape suivante est d'incorporer les deux premiers nucléotides ; puis se forme entre eux une liaison phosphodiester. Ceci crée un **complexe ternaire** polymérase-ADN-ARN en cours de synthèse. Le complexe ternaire se forme très rapidement quand l'ARN polymérase trouve un promoteur si bien que le complexe binaire a une existence extrêmement transitoire.

Le facteur sigma est seulement impliqué dans l'initiation. Il est relargué de l'enzyme « cœur » lorsque la synthèse de l'ARN a été initiée.

L'enzyme « cœur » dans le complexe ternaire est liée très fortement à l'ADN. Elle est essentiellement « verrouillée » jusqu'à ce que l'élongation soit achevée. Quand la transcription se termine, l'enzyme « cœur » est libérée de l'ADN sous forme de protéine tétramérique libre. Elle doit alors trouver un autre facteur sigma pour pouvoir entreprendre un nouveau cycle de transcription. Le rapport facteur sigma-enzyme « cœur » est de l'ordre d'un tiers.

L'ARN polymérase peut trouver des promoteurs sur l'ADN par le processus d'essai-erreur illustré Figure 9.5.

L'enzyme « cœur » en excès existe surtout sous forme de complexes lâches fermés, car l'enzyme y entre rapidement et les quitte lentement.

En revanche, l'holoenzyme s'associe aux sites de liaison lâche, et s'en dissocie très rapidement. Ainsi, tout se passe comme si se faisait et se défaisait une série de liaisons lâches d'une manière désordonnée, jusqu'à ce que (par chance) elle rencontre un promoteur. Sa reconnaissance de la séquence spécifique permettra alors à une liaison serrée d'avoir lieu par la formation d'un complexe ouvert.

Trois étapes sont nécessaires à l'ARN polymérase pour se déplacer d'un site de liaison à un autre sur l'ADN. Elle doit se dissocier du premier site de liaison, trouver le second site, et s'y associer. Le mouvement d'un site à l'autre est limité par la vitesse de diffusion dans le milieu.

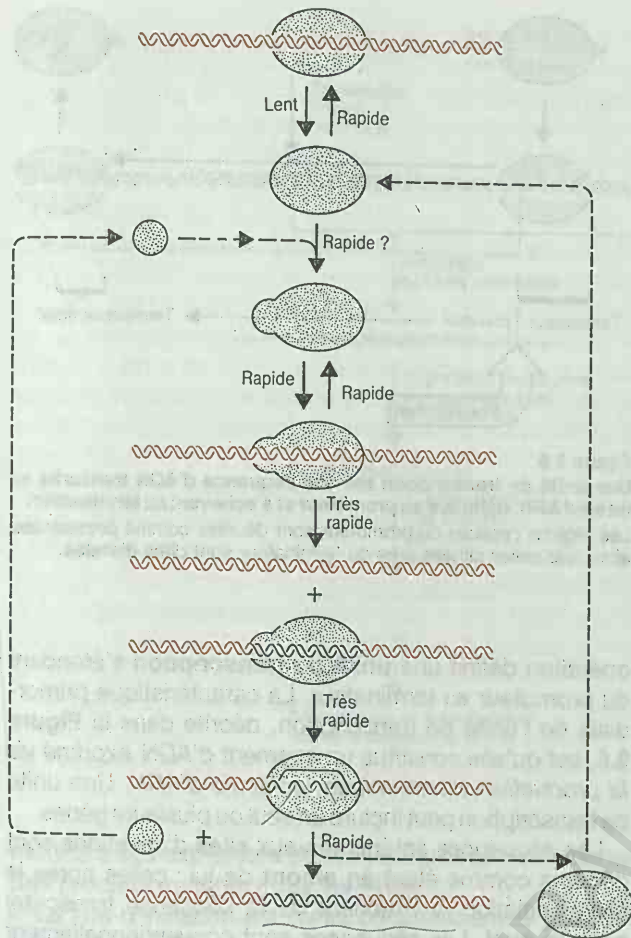


Figure 9.5
Le facteur sigma et l'enzyme « cœur » se recyclent en différents points au cours de la transcription.

Le facteur sigma est libéré aussitôt qu'un complexe ternaire s'est formé à un site d'initiation ; il devient disponible pour être utilisé par une autre enzyme « cœur ». L'enzyme « cœur » est libérée au site de terminaison : elle doit soit trouver un facteur sigma et former une holoenzyme capable de se lier de manière stable seulement aux promoteurs, soit se lier à des sites lâches de l'ADN.

La constante de vitesse de liaison aux promoteurs est très proche de cette limite, tellement proche en fait, qu'un temps insuffisant est laissé pour une association et une dissociation des sites de liaison lâche pendant un cycle aléatoire de recherche.

L'ARN polymérase doit donc utiliser d'autres moyens pour rechercher ses sites de liaison, probablement le déplacement direct d'une séquence liée par une autre. Au lieu de se déplacer en quittant un site de liaison et en diffusant vers un autre, l'enzyme vraisemblablement se

fixe à une séquence d'ADN, l'échange très rapidement pour une autre et continue d'échanger des séquences de manière confuse jusqu'à ce qu'un promoteur soit trouvé. L'enzyme forme alors un complexe ouvert stable, après quoi l'initiation a lieu. Le processus de recherche devient beaucoup plus rapide car l'association et la dissociation sont virtuellement simultanées, et il n'y a pas de temps perdu à permuter entre les sites.

L'existence d'un cycle dans lequel le facteur sigma et l'enzyme « cœur » s'associent seulement de manière temporaire résout le dilemme de l'ARN polymérase, réconciliant ses besoins pour l'initiation et pour l'élongation. C'est un dilemme car l'initiation réclame une liaison serrée *seulement* sur des séquences particulières (promoteurs), alors que l'élongation requiert une association étroite avec *toutes* les séquences le long desquelles l'enzyme doit se déplacer.

L'enzyme « cœur » a, pour l'ADN, une haute affinité intrinsèque, qui est augmentée par la présence d'un ARN naissant. Mais cette affinité pour des sites de liaisons lâches reste trop forte pour permettre à l'enzyme de trouver efficacement le promoteur ; les associations et dissociations impliquées dans le processus d'essai-erreur pour trouver un site de liaison serrée pourraient prendre plusieurs heures.

En réduisant la stabilité du complexe lâche, le facteur sigma permet au processus de se dérouler beaucoup plus rapidement, et, en stabilisant l'association aux sites de liaison serrée, le facteur mène la réaction de manière irréversible vers la formation d'un complexe ouvert. Pour éviter de devenir paralysée par son affinité spécifique pour le promoteur, l'enzyme relargue le facteur sigma, et ainsi retourne à une affinité large pour tout ADN, sans respect de séquence, ce qui lui permet de poursuivre la transcription.

Comment le facteur sigma transforme-t-il l'enzyme de telle sorte que le promoteur soit spécifiquement reconnu ? En tant que polypeptide indépendant, sigma ne semble pas se lier à l'ADN, mais lorsque l'holoenzyme forme un complexe de liaison serrée, σ entre en contact avec l'ADN dans la région de fusion initiale. L'incapacité du facteur sigma libre à reconnaître les séquences promoteurs peut être importante : si σ pouvait librement se lier aux promoteurs, il pourrait bloquer l'initiation de la transcription par l'holoenzyme. Nous ne savons pas quel rôle les sous-unités de l'enzyme « cœur » jouent dans la reconnaissance du promoteur ; sigma pourrait changer la conformation de l'enzyme « cœur » de telle manière que son aptitude à reconnaître l'ADN soit altérée, mais la spécificité de séquence de cet effet n'est pas claire.

LES UNITÉS DE TRANSCRIPTION S'ÉTENDENT DES PROMOTEURS AUX TERMINATEURS

L'initiation de la transcription est un point critique pour le contrôle de l'expression des gènes. Souvent, la décision d'initier ou non sur un promoteur particulier est l'étape majeure pour déterminer si un gène doit être exprimé. Qu'est-ce qui contrôle la capacité de l'ARN polymérase à initier sur un promoteur particulier ? Nous pouvons commencer par envisager les promoteurs comme appartenant à deux classes générales.

Certains promoteurs peuvent être reconnus par l'ARN polymérase seule ; dans ces cas, un promoteur accessible sera toujours transcrit. L'accessibilité d'un promoteur peut être déterminée par d'autres protéines, qui peuvent soit agir directement sur le promoteur pour bloquer l'accès à l'ARN polymérase, soit fonctionner indirectement en contrôlant la structure du génome dans la région.

D'autres promoteurs ne sont pas, par eux-mêmes, suffisants pour mener à bien la transcription ; des facteurs protéiques auxiliaires sont nécessaires pour que l'initiation puisse avoir lieu. Les facteurs protéiques additionnels agissent en reconnaissant des séquences d'ADN qui peuvent être proches, ou coïncider partiellement avec la séquence liée par l'ARN polymérase elle-même.

Comme les séquences d'ADN dont la fonction est d'être *reconnue par des protéines*, un promoteur et tous les sites de contrôle adjacents diffèrent des autres séquences dont le rôle est d'être transcrit ou traduit. L'information pour la fonction de promoteur est fournie directement par la séquence d'ADN : sa structure constitue le signal. Par contre, les régions exprimées acquièrent leur signification seulement après que l'information ait été transférée sous forme de quelque autre acide nucléique ou de protéine.

La question clé en examinant l'interaction entre une ARN polymérase et son promoteur est : comment une protéine peut-elle reconnaître une séquence spécifique d'ADN. Est-ce que l'enzyme possède un site actif pouvant distinguer la structure chimique d'une séquence particulière de bases dans la double hélice d'ADN ? Quel degré de spécificité doivent avoir ces exigences ? Les promoteurs varient dans leurs affinités pour l'ARN polymérase ; ceci peut être un facteur important pour le contrôle de la fréquence d'initiation et donc pour le niveau de l'expression génique.

La fixation au promoteur est rapidement suivie par l'initiation au site de départ de la transcription ; l'ARN polymérase continue alors le long de la matrice jusqu'à ce qu'elle atteigne une séquence de terminaison. Cette

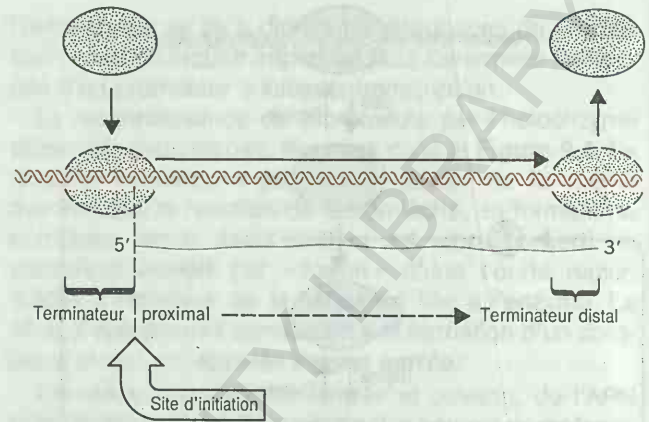


Figure 9.6

Une unité de transcription est une séquence d'ADN transcrite en un seul ARN, débutant au promoteur et s'achevant au terminateur. Les régions proches du promoteur sont décrites comme proximales, alors que celles situées près du terminateur sont dites distales.

opération définit une **unité de transcription** s'étendant du promoteur au terminateur. La caractéristique primordiale de l'unité de transcription, décrite dans la **Figure 9.6**, est qu'elle constitue un segment d'ADN *exprimé via la production d'une unique molécule d'ARN*. Une unité de transcription peut inclure un seul ou plusieurs gènes.

Les séquences antérieures aux sites d'initiations sont décrites comme étant en **amont** de lui ; celles après le site d'initiation (à l'intérieur de la séquence transcrite) sont en **aval**. Les séquences sont conventionnellement écrites de manière à ce que la transcription ait lieu de gauche (amont) à droite (aval). Ceci revient à écrire l'ARNm dans la direction usuelle de 5' vers 3'.

Souvent, la séquence de l'ADN est écrite seulement pour montrer le brin dont la séquence est la même que celle de l'ARN. La position des bases est numérotée dans les deux sens, à partir du site d'initiation, auquel on assigne la valeur +1 ; les nombres ont des valeurs croissantes en allant vers l'aval. La base précédant le site d'initiation a la valeur -1, et les nombres négatifs augmentent en se dirigeant vers l'amont.

Le produit de transcription immédiat est appelé le **transcrit primaire**. Il consisterait en un ARN s'étendant du promoteur au terminateur et qui possédant ses propres extrémités 5' et 3'. Cependant, le transcrit primaire est presque toujours instable et donc très difficile à caractériser *in vivo*. Chez les procaryotes, il est rapidement dégradé (ARNm) ou clivé pour donner des produits matures (ARNr et ARNt). Chez les eucaryotes, il est

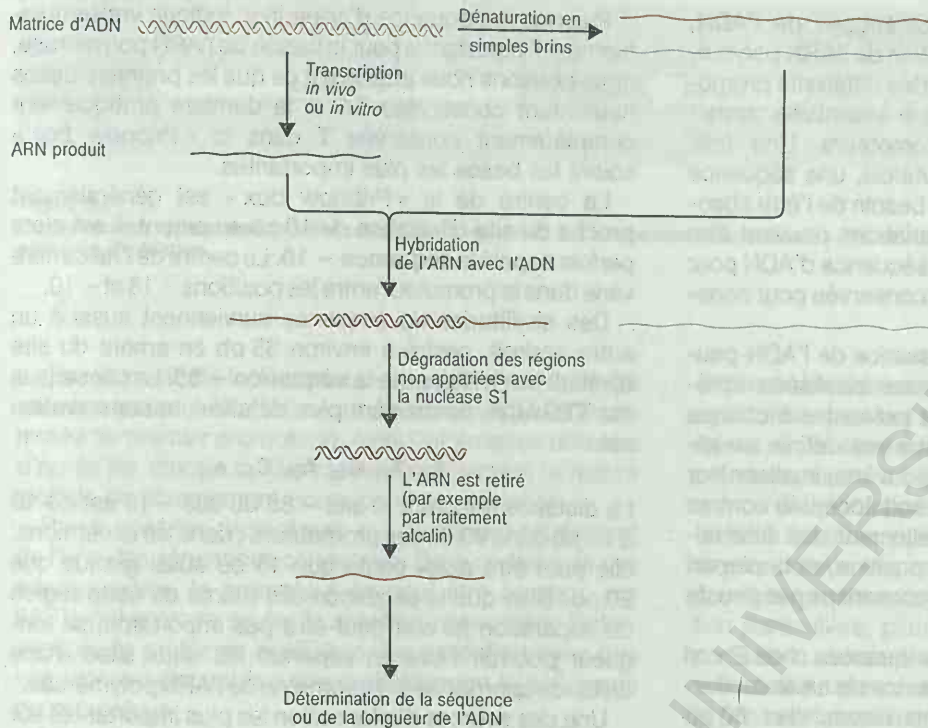


Figure 9.7
Les sites d'initiations peuvent être déterminés en comparant l'ARN produit avec l'ADN matriciel.

Lorsque l'ARN est hybridé avec les brins de l'ADN dénaturé, il forme un hybride ARN-ADN avec le brin qui lui a servi de matrice, l'autre brin d'ADN reste non apparié. Un traitement à la nucléase S1, qui dégrade spécifiquement l'ADN simple-brin, détruit à la fois les brins non appariés et les régions du brin matriciel au-delà de l'unité de transcription. L'ARN de l'hybride ARN-ADN peut être retiré. La séquence de l'ADN peut alors être déterminée ou bien il peut être utilisé pour localiser la position de l'ARN sur la matrice.

modifié à ses extrémités (ARNm) et/ou clivé pour donner des produits matures (tous les ARN).

Le site d'initiation est défini comme étant la paire de bases de l'ADN correspondant au premier nucléotide incorporé dans l'ARN. La **Figure 9.7** illustre une méthode courante pour identifier le site d'initiation en hybridant le transcrit avec sa matrice d'ADN. En principe, cela consiste à dégrader tout l'ADN qui ne peut s'hybrider avec l'ARN, et à déterminer alors la séquence de l'ADN restant.

Pour définir le site d'initiation, il est nécessaire d'examiner l'extrémité 5' du transcrit primaire. A cause de la difficulté d'isoler les transcrits primaires *in vivo*, la plupart des informations sur les sites d'initiation sont fournies par des études *in vitro*. Cependant, dans les cas où la véritable extrémité 5' a été identifiée *in vivo* (grâce à son triphosphate terminal), celle-ci est identique à l'extrémité 5' générée par transcription *in vitro*. En particulier, l'extrémité coiffée des ARNm d'eucaryotes apparaît coïncider avec le site d'initiation.

Habituellement, le site d'initiation est une unique paire de bases ; parfois il consiste en l'uné ou l'autre de deux paires de bases adjacentes, et occasionnellement il peut

impliquer n'importe quelle position adjacente. Il représente toujours un site défini du promoteur, généralement à l'intérieur de la séquence liée par l'ARN polymérase, et dans certains cas, à une extrémité.

LES PROMOTEURS COMPORTENT DES SÉQUENCES CONSENSUS

Un promoteur pourrait être défini comme une séquence invariable d'ADN reconnue par l'ARN polymérase. Dans le génome bactérien, la longueur minimale pouvant fournir un signal adéquat est 12 pb. (Toute séquence plus courte a de fortes probabilités d'apparaître — par hasard — fréquemment dans le génome, constituant des signaux erronés.) Les 12 pb de la séquence n'ont pas besoin d'être contiguës ; et en fait, si un nombre spécifique de paires de bases sépare deux séquences constantes plus courtes, leurs longueurs combinées peuvent représenter moins de 12 pb, puisque l'intervalle lui-même fournit une partie du signal (même si la séquence intermédiaire elle-même n'a pas de sens).

Essayons d'identifier les caractéristiques de l'ADN, nécessaires pour déclencher la fixation de l'ARN polymérase, en comparant les séquences des différents promoteurs. Toute séquence nucléotidique essentielle devrait être présente dans tous les promoteurs. Une telle séquence est dite **conservée**. Toutefois, une séquence conservée n'a pas nécessairement besoin de l'être absolument en tout point ; quelques variations peuvent être autorisées. Comment analyser une séquence d'ADN pour déterminer si elle est suffisamment conservée pour constituer un signal identifiable ?

Les sites présumés de reconnaissance de l'ADN peuvent être définis en tant que séquences idéalisées représentant les bases les plus souvent présentes à chaque position. Une **séquence consensus** est définie en alignant tous les exemples connus pour maximaliser leur homologie. Pour qu'une séquence soit acceptée comme consensus, chaque base individuellement doit être raisonnablement prédominante à sa position, et la plupart des séquences doivent différer du consensus par peu de substitutions, disons 1 ou 2.

Plus de 100 promoteurs ont été séquencés chez *E. coli* et il est frappant de constater l'*absence de toute conservation extensive de séquence* au niveau des 60 pb associées à l'ARN polymérase. Les séquences de la plupart des sites de liaisons peuvent en fait n'avoir aucun rapport. Mais quelques courts segments à l'intérieur du promoteur sont conservés.

Le site d'initiation est habituellement une purine (dans plus de 90 % des cas). Il est plutôt commun pour un site d'initiation d'être la base centrale de la séquence CAT, mais la conservation de ce triplet n'est pas assez importante pour le considérer comme un signal obligatoire.

Juste en amont du site d'initiation, une région de 6 pb est reconnaissable dans pratiquement tous les promoteurs. La séquence consensus est **TATAAT** ; elle est parfois appelée la boîte Pribnow (« **Pribnow box** »). La conservation de base à chaque position de la « Pribnow box » varie entre 45 % et 100 %. Le consensus peut être résumé sous la forme

$T_{80} A_{95} t_{45} A_{60} a_{50} T_{96}$

dans laquelle l'indice indique le pourcentage de présence de la base la plus fréquemment rencontrée.

(Les lettres capitales indiquent les bases conservées à plus de 54 % ; les lettres minuscules désignent les bases moins bien conservées, mais néanmoins présentes plus souvent qu'une distribution aléatoire ne pourrait le laisser supposer. Une position pour laquelle il n'y a pas de préférence discernable pour aucune des bases, serait indiquée par N.)

Puisque la fréquence d'apparition indique vraisemblablement l'importance pour la liaison de l'ARN polymérase, nous pouvons nous attendre à ce que les premières bases hautement conservées TA et la dernière pratiquement complètement conservée T dans la « Pribnow box » soient les bases les plus importantes.

Le centre de la « Pribnow box » est généralement proche du site d'initiation de 10 pb en amont. Il est alors parfois appelé la **séquence - 10**. Le centre de l'hexamère varie dans le promoteur entre les positions - 18 et - 10.

Des similitudes de séquence surviennent aussi à un autre endroit, centré à environ 35 pb en amont du site d'initiation. On l'appelle la **séquence - 35**. Le consensus est **TTGACA**, de manière plus détaillée, la conservation est

$T_{82} T_{84} G_{78} A_{65} C_{54} a_{45}$

La distance séparant le site - 35 du site - 10 est de 16 à 18 pb dans 90 % des promoteurs ; dans les exceptions, elle peut être aussi petite que 15 ou aussi grande que 20 pb. Bien que la séquence elle-même de cette région de séparation ne soit peut-être pas importante, sa longueur pourrait l'être en séparant les deux sites d'une distance appropriée à la géométrie de l'ARN polymérase.

Une des sources d'information les plus importantes sur la fonction du promoteur est fournie par les mutations. Elle affectent le niveau d'expression de(s) gène(s) que ces promoteurs contrôlent, sans pour autant altérer les produits des gènes eux-mêmes. La plupart sont identifiées sous la forme de mutants bactériens qui ont perdu, ou ont beaucoup réduit, leur faculté à transcrire les gènes adjacents. Elles sont connues comme **mutations « down »** (basses). Moins fréquemment, on trouve des mutants chez lesquels il y a une augmentation de transcription à partir du promoteur. Elles sont appelées **mutations « up »** (hautes).

Presque toutes les mutations ponctuelles affectant la fonction du promoteur tombent dans les deux séquences consensus. Occasionnellement, une mutation est localisée juste en amont de l'une ou l'autre des séquences consensus (voir Figure 9.11). Les bases présentes à d'autres positions à proximité du consensus sont nettement moins importantes ou même sont sans rôle dans la plupart des promoteurs. En plus de ces mutations, des délétions ou des insertions entre les séquences consensus peuvent altérer la distance les séparant.

Il est important de se souvenir que les mutations « up » et « down » sont définies par rapport à l'efficacité *habituelle* avec laquelle un promoteur particulier fonctionne. Ceci varie largement. Ainsi un changement reconnu comme une mutation « down » dans un promoteur peut

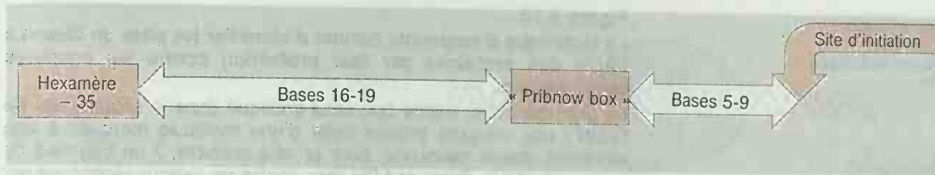


Figure 9.8

Un promoteur type possède trois composants, qui sont les séquences consensus à -35 et -10, et le site d'initiation

ne jamais avoir été isolé dans un autre (qui dans son type sauvage pourrait être même moins efficace que la forme mutée du premier promoteur). Ainsi l'information obtenue d'après les études *in vivo* identifie simplement la nature globale du changement occasionné par mutation.

Rares sont les promoteurs qui manquent totalement de l'une des séquences consensus. Dans certains de ces cas au moins, le promoteur ne peut être reconnu par l'ARN polymérase seule, la réaction nécessitant l'intervention de protéines auxiliaires. Leur réaction avec des séquences adjacentes pallie probablement la déficience du promoteur.

Nous ne connaissons pas encore tous les détails de la réaction de reconnaissance ; tout ce que l'on peut affirmer maintenant est que le promoteur « type » peut utiliser les séquences -35 et -10 pour être reconnu par l'ARN polymérase. Leur absence de promoteurs exceptionnels nous fait prendre conscience que d'autres moyens peuvent également être utilisés pour la reconnaissance, mais nous ignorons combien il existe d'alternatives, et comment exactement, elles pallient l'absence de séquences consensus.

Le promoteur le plus efficace est-il celui qui possède les séquences consensus elles-mêmes ? Cette hypothèse découle de cette simple règle qu'une mutation « up » augmente habituellement l'homologie avec l'une des séquences consensus ou les place à environ 17 pb l'une de l'autre. Les mutations « down » habituellement diminuent la ressemblance de chaque site avec les consensus ou allongent la distance entre eux à plus de 17 pb. Ces mutations ont tendance à être situées aux positions les plus hautement conservées, ce qui confirme leur importance particulière en tant que principal déterminant de l'efficacité du promoteur.

Des exceptions occasionnelles à ces règles démontrent que l'efficacité du promoteur ne peut être entièrement prévue à partir de son homologie avec les consensus. Virtuellement, tous les promoteurs actuellement connus

diffèrent des consensus ; ainsi les voisins de toute base particulière peuvent différer de promoteur à promoteur, même si la base elle-même est conservée. Nous ne pouvons pas prévoir les effets du contexte ; il est possible qu'une base n'appartenant pas au consensus à une certaine position puisse fonctionner efficacement dans un promoteur et pas dans un autre. Il est également possible que l'*exclusion* d'une certaine base d'une position particulière, plutôt que la présence d'un nucléotide particulier, puisse être un caractère important.

Toutefois, avec ces oppositions, nous pouvons définir le promoteur optimal comme une séquence constituée de l'hexamère -35, séparée par 17 pb de l'hexamère -10, s'étendant à 7 pb en amont du site d'initiation. La structure d'un promoteur optimal est illustrée Figure 9.8.

L'ARN POLYMÉRASE PEUT SE LIER AUX PROMOTEURS *IN VITRO*

Les sites de liaisons serrées, sur lesquels l'ARN polymérase forme des complexes d'initiation stables, se trouvent à l'intérieur des promoteurs. Leurs séquences peuvent être élucidées en suivant le protocole résumé dans la Figure 9.9, dans lequel l'ARN polymérase est liée *in vitro* à un fragment d'ADN contenant un promoteur. Une digestion avec l'enzyme ADNase est alors utilisée pour dégrader toutes les régions de l'ADN, non protégées par l'ARN polymérase.

Le fragment ainsi déterminé fait environ 44 pb de longueur. La polymérase peut initier la transcription si elle reste attachée au fragment protégé, et synthétiser un ARN court d'environ 20 bases, qui se termine lorsque la polymérase « s'éjecte » (run off) du fragment. Ceci montre que le fragment protégé correspond à une séquence sur laquelle la transcription est initiée, et localise le site d'initiation au centre du site de fixation. Le fragment protégé s'étend d'à peu près -20 en amont jusqu'à +20 en aval. Toutefois, l'ARN polymérase ne peut pas se

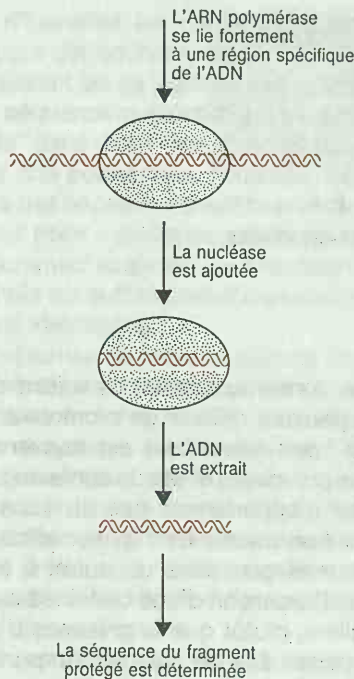


Figure 9.9

Les sites de liaison de l'ARN polymérase sur l'ADN peuvent être purifiés car ils sont protégés par l'enzyme contre la dégradation par les nucléases.

refixer aux fragments protégés (-20 à +20), des séquences additionnelles doivent donc être nécessaires à la liaison initiale.

L'aptitude de l'ADN polymérase à reconnaître l'ADN peut être caractérisée plus précisément par la **technique d'empreinte** (foot print). Une séquence d'ADN liée à l'ARN polymérase (ou à toute autre protéine) est *partiellement* digérée par une endonucléase, une enzyme qui attaque des liaisons phosphodiester à l'intérieur d'un acide nucléique. Dans des conditions appropriées, toute liaison phosphodiester qui est en principe accessible à la nucléase est coupée dans certaines molécules d'ADN, mais pas dans toutes. Une liaison donnée ne pourra absolument pas être coupée uniquement dans le cas où l'ARN polymérase bloque l'accès de l'ADN à la nucléase.

Les positions clivées sont identifiées en utilisant un ADN marqué sur un brin et à une extrémité seulement. Comme le montre la **Figure 9.10**, après le traitement à la nucléase, les fragments d'ADN digérés sont récupérés et passés sur gel d'électrophorèse, ce qui les sépare en fonction de leur longueur. Pour chaque liaison sensible

Figure 9.10

La technique d'empreinte permet d'identifier les sites de liaison à l'ADN des protéines par leur protection contre les coupures enzymatiques.

Le principe est le même que celui impliqué dans le séquençage de l'ADN ; une coupure préférentielle d'une molécule marquée à une extrémité donne naissance, pour un site sensible, à un fragment de longueur unique. Dans un ADN libre, toutes les liaisons reconnues par l'enzyme sont clivées sur l'une ou l'autre molécule. Mais lorsque l'ADN est complexé à une protéine, la région couverte par la protéine liant l'ADN est protégée dans chaque molécule. Deux réactions sont donc menées en parallèle sur : un mélange témoin ne contenant que de l'ADN pur, et un mélange expérimental contenant la protéine.

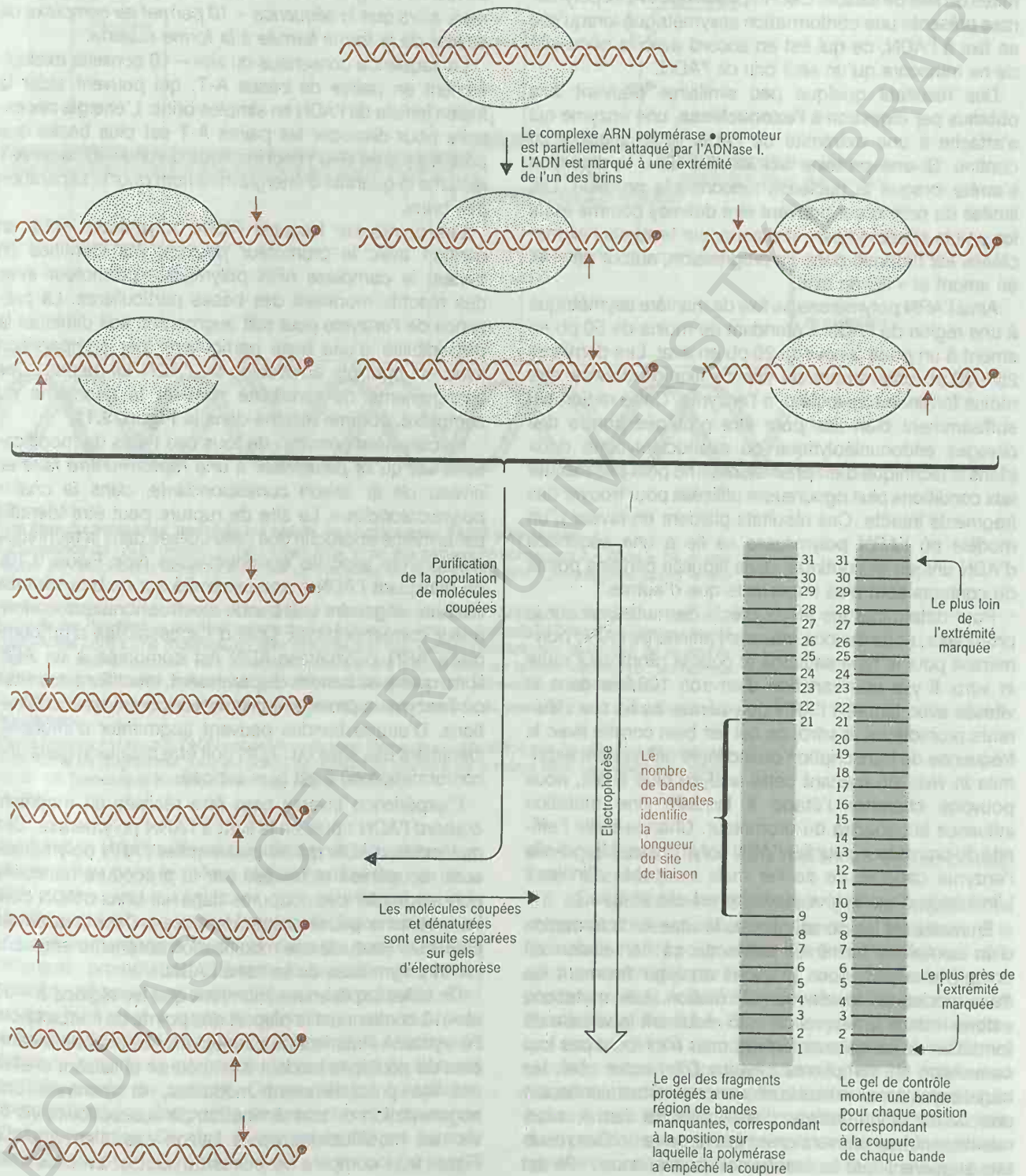
Lorsque les brins sont séparés et passés sur électrophorèse, une bande radioactive est produite par chaque fragment possédant une extrémité marquée. La position de la bande correspond au nombre de bases dans le fragment. Les plus courts migrent les plus vite ; ainsi la distance séparant la coupure de l'extrémité marquée est mesurée à partir du bas du gel (voir Figure 4.12). Dans le témoin, toutes les liaisons sont clivées, ce qui engendre une série de bandes, représentant chacune une base. Dans la figure, on peut compter 31 bandes. Dans le fragment protégé, les liaisons ne peuvent être clivées dans la région liée par la protéine, donc les bandes représentant les fragments de tailles correspondantes ne sont pas obtenues. L'absence de bandes 10-20 dans la figure identifie un site de liaison de la protéine couvrant la région située entre 10 et 20 bases de l'extrémité marquée de l'ADN.

Au lieu d'utiliser un seul gel qui analyse l'ADN simplement en fonction de la longueur, le contrôle et les mélanges expérimentaux peuvent être traités pour donner quatre gels de séquence (voir Figure 4.13). La comparaison des deux types de gel permet de lire directement la séquence nucléotidique et ainsi d'identifier celle du site de liaison.

à l'enzyme et suivant sa position, une bande est présente sur le gel ; elle correspond à la distance séparant le site de coupure de l'extrémité marquée. Pour chaque position protégée contre le clivage par l'ARN polymérase, une bande est manquante. Chacun des deux brins de l'ADN peut être analysé séparément en l'utilisant comme brin marqué. En combinant la technique d'empreinte avec le séquençage d'ADN, la séquence nucléotidique du site de liaison peut être déterminée de même que sa position.

Analysé par cette technique, le site de fixation de l'ARN polymérase apparaît plus étendu que lorsqu'il est obtenu par l'analyse des fragments protégés. Il se prolonge approximativement de 30 pb supplémentaires en amont, ce qui fait à peu près de -50 à +20. Cette zone inclut toutes les séquences nécessaires à la liaison et à l'initiation. La séquence consensus -10 se trouve au centre de la région fortement protégée ; la séquence consensus -35 se situe à une extrémité du site de liaison, mais n'est pas fortement protégée.

Les deux brins d'ADN ne sont pas aussi bien protégés dans toutes les régions du promoteur, surtout aux extré-



mités du site de liaison. Ceci implique que l'ARN polymérase présente une conformation assymétrique lorsqu'elle se fixe à l'ADN, ce qui est en accord avec la nécessité de ne transcrire qu'un seul brin de l'ADN.

Des résultats quelque peu similaires peuvent être obtenus par digestion à l'**exonucléase**, une enzyme qui s'attache à une extrémité de l'ADN et le dégrade en continu. Si une protéine est liée à l'ADN, la digestion s'arrête lorsque la nucléase rencontre la protéine. Les limites du promoteur peuvent être définies comme étant les points situés de part et d'autre, sur lesquels l'exonucléase est bloquée dans sa progression, autour de - 44 en amont et + 20 en aval.

Ainsi l'ARN polymérase se fixe de manière asymétrique à une région de l'ADN s'étendant de moins de 50 pb en amont à un point à environ 20 pb en aval. Les dernières 25 - 30 pb côté amont du site de fixation doivent être moins fortement associées à l'enzyme. Cette région est suffisamment bien liée pour être protégée contre des clivages endonucléolytique ou exonucléolytique doux (dans la technique d'empreinte), mais ne peut pas résister aux conditions plus rigoureuses utilisées pour trouver des fragments intacts. Ces résultats plaident en faveur d'un modèle où l'ARN polymérase se lie à une séquence d'ADN unique et continue, dans laquelle certains points de contacts sont plus importants que d'autres.

Pour déterminer les effets précis de mutations sur le promoteur, nous devons mesurer l'affinité de l'ARN polymérase pour le type sauvage et pour le promoteur muté *in vitro*. Il y a une variation d'environ 100 fois dans la vitesse avec laquelle l'ARN polymérase se lie aux différents promoteurs *in vitro*, ce qui est bien corrélé avec la fréquence de transcription quand leurs gènes sont exprimés *in vivo*. En menant cette analyse plus avant, nous pouvons chercher l'étape à laquelle une mutation influence la capacité du promoteur. Change-t-elle l'affinité du promoteur pour lier l'ARN polymérase ? Rend-elle l'enzyme capable de se lier mais incapable d'initier ? L'influence d'un facteur auxiliaire est-elle altérée ?

En mesurant les constantes de la vitesse de formation d'un complexe fermé et celles de sa conversion en complexe ouvert, nous pouvons analyser finement les deux stades de la réaction d'initiation. Les mutations « down » dans la séquence - 35 réduisent la vitesse de formation de complexes fermés, mais n'inhibent pas leur conversion en complexes ouverts. D'un autre côté, les mutations « down » dans la séquence - 10 ne ralentissent pas la formation initiale d'un complexe fermé, mais ralentissent leur conversion en forme ouverte. Ces résultats suggèrent que la *fonction de la séquence - 35 est*

de fournir le signal de reconnaissance par l'ARN polymérase, alors que la séquence - 10 permet au complexe de passer de la forme fermée à la forme ouverte.

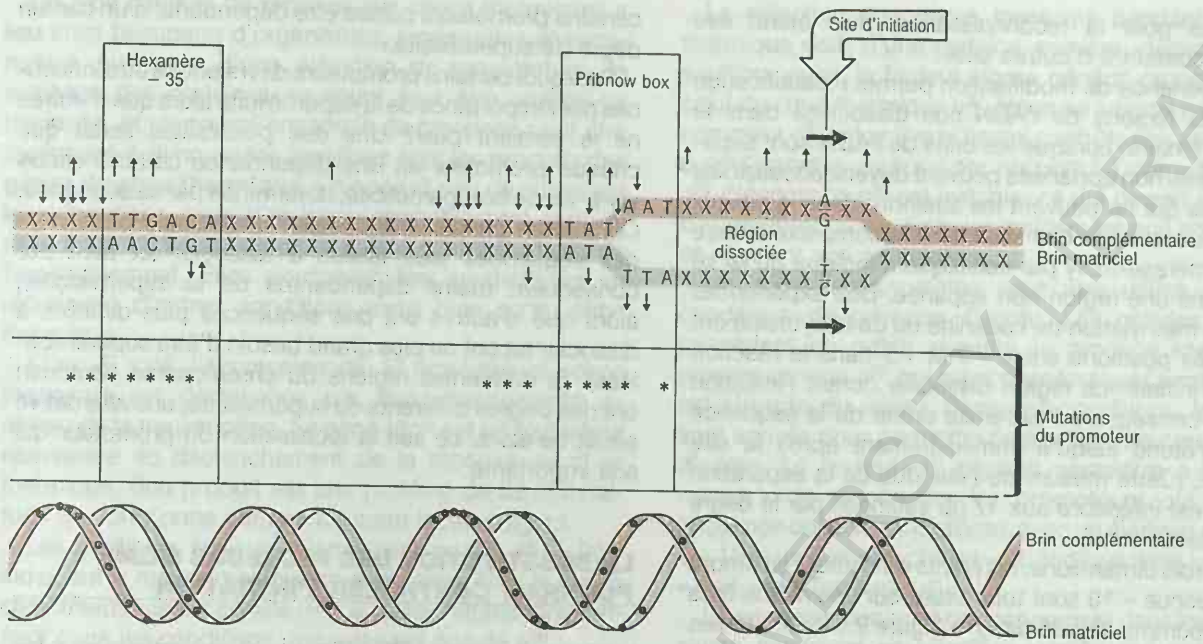
La séquence consensus du site - 10 consiste exclusivement en paires de bases A-T, qui peuvent aider la fusion initiale de l'ADN en simples brins. L'énergie nécessaire pour dissocier les paires A-T est plus basse que pour les paires G-C ; ceci implique qu'une séquence A-T réclame la quantité d'énergie minimum pour la séparation des brins.

Les points sur lesquels l'ARN polymérase entre en contact avec le promoteur peuvent être identifiés en traitant le complexe ARN polymérase-promoteur avec des réactifs modifiant des bases particulières. La présence de l'enzyme peut soit augmenter, soit diminuer la disponibilité d'une base particulière (par comparaison avec un contrôle consistant en l'ADN lui-même). Ces changements de sensibilité révèlent la géométrie du complexe, comme résumé dans la **Figure 9.11**.

Le caractère commun de tous ces types de modifications est qu'ils *permettent à une rupture d'être faite au niveau de la liaison correspondante, dans la chaîne polynucléotidique*. Le site de rupture peut être identifié par la même approche que celle utilisée dans la technique d'empreinte avec les endonucléases (voir Figure 9.10). En marquant l'ADN à une extrémité sur un brin, chaque coupure engendre une bande électrophorétique de longueur correspondante. Quand l'accessibilité d'un complexe ARN polymérase-ADN est comparée à un ADN libre, quelques bandes disparaissent, identifiant des sites où l'enzyme a protégé le promoteur contre les modifications. D'autres bandes peuvent augmenter d'intensité, identifiant des sites où l'ADN doit être maintenu dans une conformation où il est plus exposé.

L'expérience inverse peut être réalisée en modifiant *d'abord* l'ADN ; puis en le liant à l'ARN polymérase. Ces molécules d'ADN qui ne peuvent lier l'ARN polymérase sont récupérées et traitées par la procédure habituelle pour engendrer des coupures dans les brins d'ADN dont les positions peuvent être identifiées. Ceci localise les points sur lesquels une modification antérieure *empêche* l'ARN polymérase de se lier à l'ADN.

De telles expériences montrent que les régions à - 35 et - 10 contiennent la plupart des points de contact pour l'enzyme. A l'intérieur de ces régions, les mêmes ensembles de positions tendent à empêcher la liaison, si elles ont été précédemment modifiées, et montrent une augmentation ou une diminution de susceptibilité vis-à-vis des modifications après liaison avec l'enzyme. La Figure 9.11 compare les points de contact avec les sites

**Figure 9.11**

Une face du promoteur possède les points de contact pour l'ARN polymérase

La séquence de l'ADN montre un promoteur type, avec des séquences consensus à - 35 et - 10, et une région de dissociation initiale s'étendant de l'intérieur de la « Pribnow box » jusque juste après le site d'initiation.

Partie supérieure : Les sites sur lesquels une modification empêche l'ARN polymérase de se lier, sont indiqués par des flèches pointées vers la double hélice. Les sites sur lesquels l'ADN est protégé par l'ARN polymérase contre une modification sont indiqués par des flèches pointant vers l'extérieur de la double hélice. Les flèches pointant vers les bases indiquent une modification sur la base elle-même ; les flèches pointant entre les bases indiquent une modification de la liaison de connexion phosphodiester.

Au centre : Les sites sur lesquels des mutations affectent la fonction du promoteur sont indiqués par des astérisques.

Partie inférieure : Lorsque l'ADN est représenté sous forme de double hélice vue de côté, comme indiqué sous forme de diagramme, tous les points de contact se trouvent sur la même face. La plupart sont situés sur le brin homologue (celui qui n'est pas le brin matriciel).

de mutation ; bien qu'ils ne coïncident pas complètement, elles ont lieu dans la même région limitée. Sur les deux sites consensus, la région de contact s'étend sur 12-15 pb, ce qui est un peu plus long que la région conservée.

Il est remarquable que les mêmes positions dans différents promoteurs puissent fournir des points de contact, même si une base différente est présente. Ceci indique qu'il peut y avoir un mécanisme commun pour la fixation de l'ARN polymérase, bien que la réaction ne dépende pas de la présence de bases particulières à certains points de contact. Ce modèle peut expliquer pourquoi certains des points de contact ne sont pas des sites de mutation. De plus, toutes les mutations ne sont pas situées à un point de contact ; le voisinage peut-il

avoir quelque influence sans réellement être en contact avec l'enzyme ?

Il est spécialement significatif que les expériences comportant une modification des sites, antérieure à la fixation de l'ARN polymérase, identifient les mêmes sites que ceux protégés par l'enzyme contre une modification postérieure à sa fixation. Ces deux expériences mesurent des choses différentes. La première identifie tous les sites que l'enzyme doit reconnaître pour se lier à l'ADN. La seconde reconnaît tous les sites qui en fait établissent des contacts dans le complexe binaire. Les sites protégés incluent tous les sites de reconnaissance et aussi quelques positions supplémentaires ; ceci suggère que l'enzyme reconnaît en premier un ensemble de sites lui étant

nécessaires pour la reconnaissance, puis étend ses points de contact à d'autres sites.

Une expérience de modification permet l'identification directe des régions de l'ADN non dissociées dans le complexe binaire. Lorsque les brins de l'ADN sont séparés, les bases non appariées peuvent devenir accessibles aux réactifs qui ne peuvent les atteindre dans la double hélice. L'accessibilité des sites dans le complexe binaire ARN polymérase-ADN par conséquent indique qu'ils se situent dans une région non appariée. Des expériences utilisant la méthylation de l'adénine ou de la cytosine ont détecté des positions entre - 9 et + 3 dans la réaction de fusion initiale. La région dissociée durant l'initiation inclut par conséquent l'extrémité droite de la séquence - 10 et s'étend jusqu'à immédiatement après le site d'initiation. (Cette mesure de l'étendue de la séparation des brins est inférieure aux 17 pb estimées par le degré total de dissociation).

Vus en trois dimensions, les points de contact en amont de la séquence - 10 sont tous situés sur une même face de l'ADN comme illustré dans la Figure 9.11. Ces bases peuvent être reconnues lors de la formation initiale d'un complexe binaire fermé. Ceci pourra rendre possibles l'approche et la reconnaissance par l'ARN polymérase d'une face de l'ADN. Alors que la dissociation de l'ADN commence, d'autres sites, qui, à l'origine, étaient situés sur l'autre face de l'ADN, pourraient être reconnus et liés.

L'importance de la séparation des brins dans l'initiation de la transcription est soulignée par les effets de la superhélicité. Les ARN polymérases de procaryotes et d'eucaryotes peuvent toutes deux initier la transcription plus efficacement *in vitro* lorsque la matrice est superenroulée, probablement parce que cette structure requiert moins d'énergie libre pour la fusion initiale de l'ADN dans le complexe d'initiation.

Le rôle de ce phénomène dans le contrôle de l'activité des promoteurs bactériens est souligné par l'influence des enzymes qui modifient le degré de superhélicité. Parmi ces enzymes, on trouve l'ADN gyrase, qui *introduit* des supertours négatifs, et la topo-isomérase I, qui *relâche* les supertours négatifs (voir Chapitre 28). Les inhibiteurs de l'ADN gyrase réduisent la transcription ; les mutations de la topo-isomérase I peuvent augmenter la transcription. Ces deux effets ne sont visibles que sur certains promoteurs.

La nature de ces effets n'est pas totalement claire ; la bactérie s'efforce évidemment de maintenir le degré de superhélicité entre certaines limites, car des mutations dans une enzyme, qui en altèrent le niveau, peuvent être compensées par des mutations dans d'autres pour restaurer l'équilibre. Toutefois, il semble que l'efficacité de

certains promoteurs puisse être dépendante d'un certain degré de superhélicité.

Pourquoi certains promoteurs devraient-ils être influencés par l'importance de la superhélicité alors que d'autres ne le seraient pas ? Une des possibilités serait que chaque promoteur ait une dépendance caractéristique vis-à-vis de la superhélicité, déterminée par sa séquence. Ceci laisse penser que certains promoteurs possèdent des séquences plus faciles à dissocier (et sont par conséquent moins dépendantes de la superhélicité), alors que d'autres ont des séquences plus difficiles à dissocier (et ont un plus grand besoin d'être superenroulées). Si différentes régions du chromosome bactérien ont des degrés différents de superhélicité, une alternative serait qu'alors, ce soit la localisation du promoteur qui soit importante.

LA SUBSTITUTION DES FACTEURS SIGMA POURRAIT CONTRÔLER L'INITIATION

L'ARN polymérase d'*E. coli* transcrit en pratique tous les gènes bactériens, et doit donc reconnaître un large spectre de promoteurs, dont les gènes associés sont transcrits à des niveaux différents dans différentes occasions. La capacité d'initier sur des promoteurs particuliers est contrôlée par des séries apparemment infinies de facteurs auxiliaires, coopérant ou interférant avec l'enzyme.

Dans pratiquement aucun de ces cas, il n'y a de changement au niveau des sous-unités de l'enzyme elle-même. Il est probable qu'un tel mécanisme ne soit pas favorisé, car un changement, qui permettrait à l'ARN polymérase de reconnaître spécifiquement un type de promoteur, pourrait l'empêcher d'en trouver un autre. Ceci réduirait l'adaptabilité de l'enzyme.

Cependant, la division du travail entre une enzyme « cœur », qui prend en charge l'élongation des chaînes, et un facteur sigma, impliqué dans la sélection des sites, pose immédiatement la question de savoir s'il pourrait y avoir plus d'un type de sigma, chacun étant spécifique d'une classe différente de promoteurs.

Des changements dans les facteurs sigma semblent habituellement avoir lieu, seulement lorsqu'il y a une réorganisation globale de la transcription. Pendant longtemps, *E. coli* fut considéré comme ne possédant qu'un seul facteur sigma. A présent, d'autres ont été découverts. L'un d'entre eux est utilisé dans les conditions de choc thermique (« heat shock »), lorsque les bactéries changent leur profil de transcription à cause d'une élévation de température.

Un type habituel de réponse aux chocs thermiques a lieu chez beaucoup d'organismes, procaryotes et eucaryotes. Au cours d'une élévation de température, la synthèse des protéines en cours peut être arrêtée ou diminuée, et un nouvel ensemble de protéines peut être synthétisé. Les nouvelles protéines sont les produits des **gènes du choc thermique**. On connaît peu de chose sur leurs fonctions, mais elles jouent probablement certains rôles dans la protection de la cellule contre le stress de l'environnement. Elles pourraient être synthétisées en réponse à d'autres conditions aussi bien qu'au choc thermique.

Chez *E. coli*, l'expression de 17 protéines du choc thermique est déclenchée par des changements au niveau de la transcription. Le gène *htpR* est un régulateur nécessaire au déclenchement de la réponse au choc thermique. Son produit est une protéine de 32 000 daltons qui fonctionne comme un autre facteur sigma.

Les multiples facteurs sigma sont notés σ^{00} , « 00 » indiquant la masse moléculaire du facteur. Ainsi le sigma choc thermique est appelé σ^{32} ; le facteur sigma fonctionnant dans les conditions normales est appelé σ^{70} .

σ^{32} conduit l'enzyme « cœur » à initier la transcription au niveau des promoteurs des gènes du choc thermique. Ces promoteurs ont-ils une séquence particulière qui les rendent identifiables par l'holoenzyme contenant σ^{32} ? Ils possèdent une séquence - 35 étendue et légèrement différente comparée au consensus habituel, mais ils ont un consensus entièrement différent à - 10. Dans le **Tableau 9.4**, les séquences consensus des promoteurs reconnus par σ^{70} sont comparées avec celles du choc thermique reconnues par σ^{32} .

Les promoteurs du choc thermique ont une séquence consensus - 35 qui possède un tétramère en commun avec la séquence - 35 des promoteurs généraux. La différence dans la reconnaissance du promoteur entre l'ARN polymérase sous le contrôle de σ^{70} et l'ARN polymérase contenant σ^{32} , caractéristique du choc thermique, peut donc résider dans la reconnaissance de la séquence consensus - 10 par le facteur sigma.

La sélection des gènes transcrits pendant le choc thermique doit, d'une certaine manière, dépendre d'un équilibre entre le facteur sigma général (appelé σ^{70}) et celui du choc thermique σ^{32} . Nous ne savons pas encore comment σ^{32} est activé ni ce qui contrôle son association avec l'enzyme « cœur » par rapport à l'interaction avec σ^{70} . Néanmoins σ^{32} est instable, ce qui devrait permettre d'augmenter ou de diminuer rapidement son abondance en régulant son expression.

Un autre facteur sigma peut être utilisé dans les conditions de privation d'azote. Les cellules d'*E. coli* possèdent en petite quantité la protéine identifiée à présent comme σ^{60} , qui est activée lorsque l'ammoniaque est absente du milieu. Dans ces conditions, les gènes sont activés pour permettre l'utilisation d'autres sources d'azote. Le facteur σ^{60} pourrait permettre à l'enzyme « cœur » de reconnaître les promoteurs qui ont une séquence consensus distincte, avec un élément conservé à - 10 et un autre proche de - 20 (indiqué dans la colonne « - 35 » du Tableau 9.4).

Un exemple plus large de changement dans les facteurs sigma est fourni par *Bacillus subtilis*, qui possède de multiples facteurs sigma avec différentes spécificités. Dans certaines circonstances, lorsqu'un changement majeur survient dans le mode de vie de la cellule, il y a une réorientation massive dans l'expression des gènes. De ce fait, il y a évidemment moins d'obstacle à introduire des changements permanents dans l'ARN polymérase elle-même, et *B. subtilis* remplace le facteur sigma initial par d'autres dont la spécificité de promoteur est différente.

La principale ARN polymérase trouvée dans les cellules de *B. subtilis* engagées dans une croissance végétative normale a la même structure que celle de *E. coli*, $\alpha_2\beta\beta'\sigma$. Le facteur sigma a une masse de 43 000 daltons et est décrit comme σ^{43} . Il reconnaît les promoteurs avec les mêmes séquences consensus que celles utilisées par l'enzyme d'*E. coli* sous le contrôle de σ^{70} . Des variants de l'enzyme contenant d'autres facteurs sigma sont présents en quantités beaucoup plus faibles. Les variants de l'enzyme reconnaissent différents promoteurs.

Tableau 9.4
Les facteurs sigma d'*E. coli* reconnaissent des promoteurs avec des séquences consensus différentes.

Facteur	Gène	Rôle	Séquence-35	Séparation	Séquence-10
σ^{70}	<i>rpoD</i>	général	TTGACA	16-18 pb	TATAAT
σ^{32}	<i>rpoH</i>	choc thermique	CNCTTGAA	13-15 pb	CCCCATNT
σ^{60}	<i>rpoN</i>	azote	CTGGNA	6 pb	TTGCA

Les gènes codant pour les facteurs sigma ont la désignation *rpo* pour la sous-unité de l'ARN polymérase, suivie d'une lettre pour identifier le gène lui-même [D pour le facteur sigma original, H pour « heat shock » (choc thermique), N pour privation d'azote].

Les substitutions de facteurs sigma, qui provoquent le passage de l'expression d'un jeu de gènes à l'expression d'un autre jeu, surviennent pendant l'infection par des bactériophages. L'**infection lytique** est le mécanisme par lequel un phage (virus de bactérie) prend le contrôle d'une bactérie et la détruit lors du processus de reproduction des particules phagiques. Dans tous les cas les plus simples, le développement du phage implique un changement dans les profils de transcription. Chez *E. coli*, ces changements sont réalisés grâce à la synthèse d'une ARN polymérase codée par le phage ou par l'action de facteurs auxiliaires codés par le phage et qui contrôlent l'ARN polymérase bactérienne. Lors de l'infection de *B. subtilis* par le phage SPO1, deux nouveaux facteurs sigma sont toutefois élaborés.

Le cycle infectieux de SPO1 passe par trois étapes de l'expression génique. Lors de l'infection, les gènes **précoces** du phage sont immédiatement transcrits. Après 4-5 minutes, les gènes précoces cessent d'être transcrits et les gènes **intermédiaires** le sont à leur tour. Puis vers 8-12 minutes, la transcription des gènes intermédiaires fait place à la transcription des gènes **tardifs**.

Les gènes précoces sont transcrits par l'holoenzyme de la bactérie hôte. Ils sont essentiellement indistinguishables des gènes de l'hôte dont les promoteurs ont la capacité intrinsèque d'être reconnus par l'ARN polymérase $\alpha_2\beta\beta'\sigma^{43}$.

L'expression des gènes du phage est nécessaire à la transcription des gènes intermédiaires et tardifs. Trois gènes de régulation, appelés 28, 33 et 34, contrôlent le déroulement de la transcription. Leurs fonctions sont résumées dans la **Figure 9.12**. Ce modèle de régulation crée une **cascade** d'événements, dans laquelle l'enzyme de l'hôte transcrit un gène précoce dont le produit est nécessaire à la transcription des gènes intermédiaires ; deux de ces gènes intermédiaires codent ensuite pour des produits nécessaires à la transcription des gènes tardifs.

Les mutants du gène précoce 28 ne peuvent pas transcrire les gènes intermédiaires. Le produit du gène 28 (appelé gp28) est une protéine de 26 000 daltons qui remplace le facteur sigma sur l'enzyme « cœur ». *Cette substitution est l'unique événement nécessaire à réaliser la transition entre l'expression de gènes précoces et celle des gènes intermédiaires.* Elle crée une enzyme complète qui ne peut plus transcrire les gènes de l'hôte, mais transcrit spécifiquement les gènes intermédiaires à leur place. Nous ignorons comment gp28 remplace σ^{43} , et ce qui arrive au polypeptide sigma de l'hôte. Il est probable que gp28 a une plus grande affinité pour l'enzyme « cœur ».

Deux des gènes intermédiaires sont impliqués dans la transition suivante. Des mutations dans l'un des gènes 33 ou 34 empêchent la transcription des gènes tardifs. Les produits de ces gènes sont respectivement des protéines de 13 000 et 24 000 daltons, qui remplacent gp28 sur l'enzyme « cœur » de la polymérase. Encore une fois, nous ignorons comment gp33 et gp34 excluent gp28 (ou tout autre σ^{43} résiduel de l'hôte), *mais une fois qu'ils sont liés à l'enzyme « cœur », elle n'est capable d'initier la transcription que sur les promoteurs des gènes tardifs.*

Les remplacements successifs du facteur sigma ont une double conséquence. Chaque fois que la sous-unité est changée, l'ARN polymérase devient capable de reconnaître une nouvelle classe de gènes, *et ne reconnaît plus la classe précédente.* Ces substitutions constituent donc des changements globaux dans l'activité de l'ARN polymérase. Probablement tous ou pratiquement tous les enzymes « cœur » se trouvent associés au facteur sigma du moment ; et le changement est irréversible.

Les nouveaux facteurs sigma sont également utilisés pendant la **sporulation**, un mode de vie alternatif dont disposent quelques bactéries. A la fin de la **phase végétative**, la croissance exponentielle cesse car le milieu s'épuise. Ceci déclenche la sporulation ; l'ADN est répliqué, un génome est isolé à une extrémité de la cellule, et est éventuellement entouré par la solide enveloppe de la spore. Le procédé prend environ 8 heures.

La sporulation implique un changement complet dans les activités de biosynthèse de la bactérie, et beaucoup de gènes sont concernés. Le niveau fondamental du contrôle se trouve au niveau de la transcription. Certains des gènes fonctionnant lors de la phase végétative sont rendus silencieux pendant la sporulation, mais la plupart continuent à être exprimés. De plus, les gènes spécifiques de la sporulation ne sont exprimés que durant cette période. A la fin de la sporulation, environ 40 % des ARNm de la bactérie sont spécifiques de la sporulation.

Dé nouvelles formes d'ARN polymérase deviennent actives dans les cellules en sporulation ; elles contiennent la même enzyme « cœur » que les cellules végétatives, mais possèdent différentes protéines à la place de la forme végétative σ^{43} . Ces changements dans la spécificité de la transcription sont résumés dans la **Figure 9.13**.

Au début de la sporulation, σ^{43} est remplacé par σ^{37} , une protéine plus petite qui, de fait, est présente dans les cellules végétatives, bien qu'alors elle ne soit pas associée avec l'enzyme « cœur ». Sous le contrôle de σ^{37} , l'ARN polymérase transcrit le premier jeu de gènes de sporulation au lieu des gènes végétatifs qu'elle transcrivait précédemment.

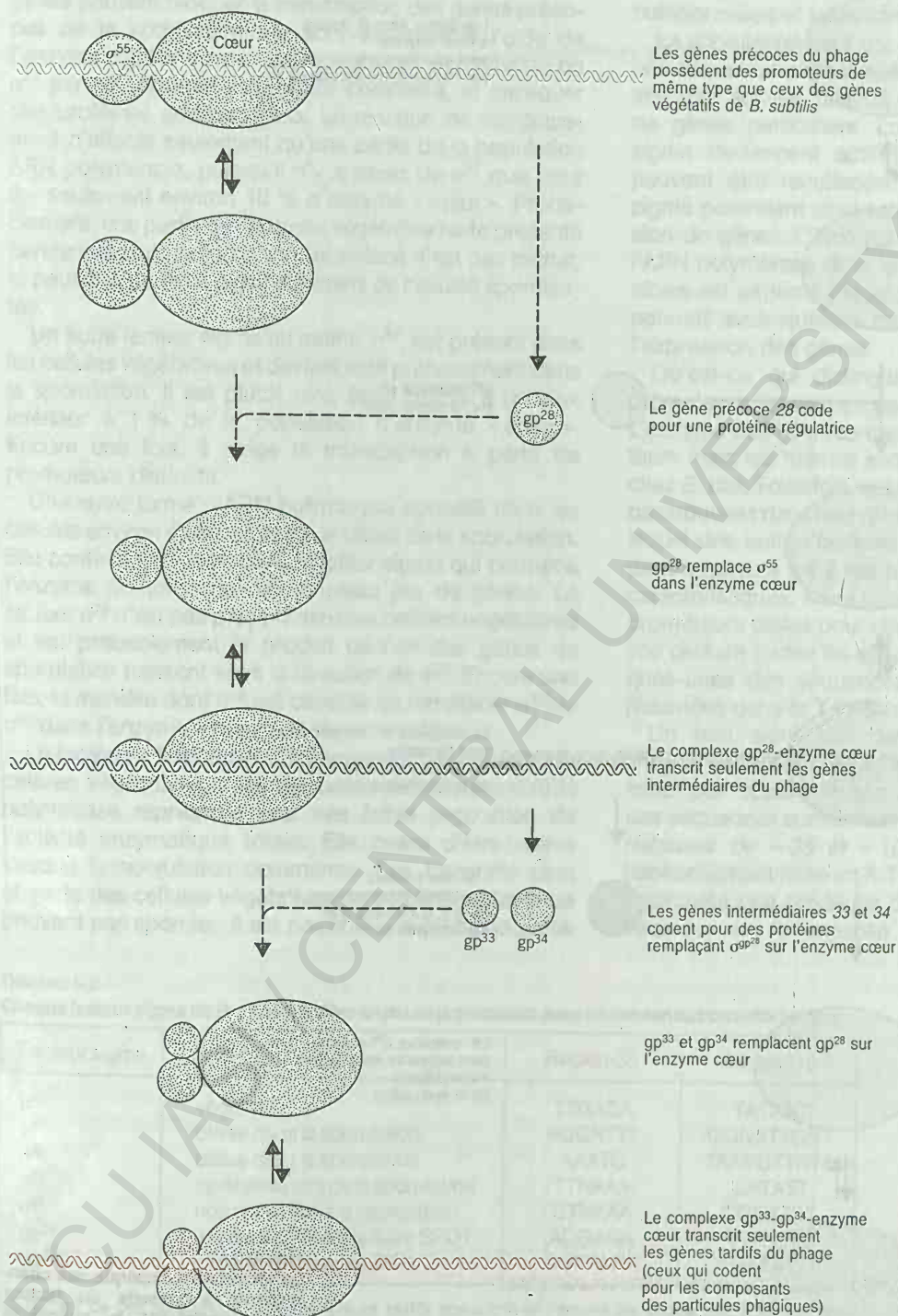


Figure 9.12
La transcription des gènes SPO1 du phage est contrôlée par deux substitutions successives du facteur sigma, ce qui change la spécificité d'initiation.

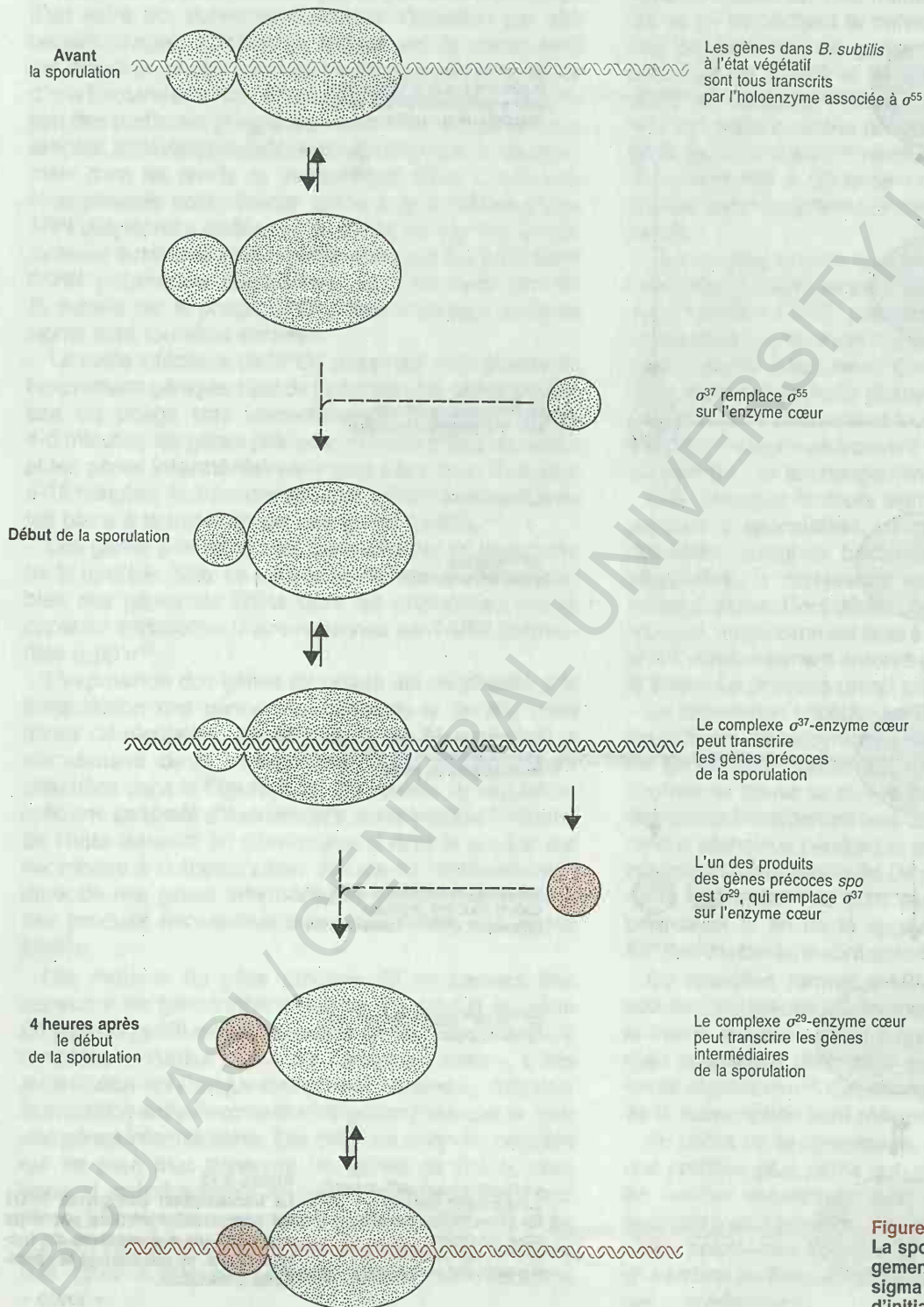


Figure 9.13

La sporulation implique des changements successifs du facteur sigma contrôlant la spécificité d'initiation de l'ARN polymérase.

Qu'est-ce qui contrôle la chronologie du remplacement ? Des mutations dans n'importe lequel des huit gènes peuvent bloquer la transcription des gènes précoces de la sporulation qui sont exprimés à l'aide de l'enzyme $\alpha_2\beta'\sigma^{43}$, ainsi le processus de substitution de σ^{43} par σ^{37} , pourrait être plutôt complexe, et impliquer des protéines additionnelles. La réaction de remplacement n'affecte seulement qu'une partie de la population ARN polymérase, puisqu'il n'y a assez de σ^{37} que pour lier seulement environ 10 % d'enzyme « cœur ». Probablement, une partie de l'enzyme végétative reste présente pendant la sporulation. Le σ^{43} remplacé n'est pas détruit, et peut être purifié à partir d'extraits de cellules sporulantes.

Un autre facteur sigma au moins, σ^{32} , est présent dans les cellules végétatives et devient actif précocement dans la sporulation. Il est plutôt rare, étant trouvé à un taux inférieur à 1 % de la population d'enzyme « cœur ». Encore une fois, il dirige la transcription à partir de promoteurs distincts.

Une autre forme d'ARN polymérase apparaît dans les cellules environ 4 heures après le début de la sporulation. Elle contient σ^{29} , un nouveau facteur sigma qui permet à l'enzyme de transcrire un nouveau jeu de gènes. Le facteur σ^{29} n'est pas présent dans les cellules végétatives et est probablement le produit de l'un des gènes de sporulation transcrit sous la direction de σ^{37} . Encore une fois, la manière dont σ^{29} est capable de remplacer σ^{37} ou σ^{43} dans l'enzyme « cœur » reste un mystère.

Le facteur σ^{28} est associé à l'enzyme « cœur » dans les cellules végétatives, dans lesquelles cette forme d'ARN polymérase représente une très faible proportion de l'activité enzymatique totale. Elle cesse d'être active lorsque la sporulation commence. Ses transcrits sont absents des cellules végétatives de certains mutants ne pouvant pas sporuler. Il est possible que σ^{28} fasse partie

de certains systèmes de signal, responsables de l'expression de gènes dont les produits détectent des carences nutritionnelles et initialisent la réponse sporulative.

La sporulation peut par conséquent être contrôlée par une séquence dans laquelle des facteurs sigma successifs sont activés, chacun dirigeant la synthèse d'un jeu de gènes particuliers. Lorsque de nouveaux facteurs sigma deviennent actifs, les anciens facteurs sigma peuvent être remplacés ; les transitions des facteurs sigma pourraient ainsi activer ou bien éteindre l'expression de gènes. L'incorporation de chaque facteur dans l'ARN polymérase dicte le moment où son jeu de gènes cibles est exprimé ; et la quantité du facteur disponible pourrait avoir quelque chose à voir avec le niveau de l'expression des gènes.

Qu'est-ce qui distingue les différentes classes de promoteurs reconnues par les divers facteurs sigma ? L'enzyme de l'hôte, contenant σ^{43} , reconnaît des promoteurs avec les mêmes séquences - 35 et - 10 décrites chez *E. coli*. Toutefois, ces séquences consensus ne sont pas trouvées dans les promoteurs reconnus par n'importe lequel des autres facteurs sigma. Peut-être chaque jeu de promoteurs a-t-il ses propres séquences consensus caractéristiques. Nous n'avons pas caractérisé assez de promoteurs cibles pour chaque facteur sigma pour pouvoir déduire toutes les séquences consensus, mais quelques-unes des séquences conservées possibles sont résumées dans le **Tableau 9.5**.

Un trait significatif des promoteurs pour chaque enzyme est qu'ils ont même taille et même emplacement relatif par rapport au site d'initiation, et qu'ils montrent des séquences conservées seulement autour des centres habituels de - 35 et - 10. Le consensus à - 10 est habituellement riche en A-T ; le consensus à - 35 montre également une tendance dans ce sens. Nous ignorons l'étendue de la capacité d'utiliser différents facteurs

Tableau 9.5

Chaque facteur sigma de *B. subtilis* utilise un jeu de promoteurs avec un consensus caractéristique.

Facteur sigma	Source et usage	Région-35	Région-10
σ^{43}	végétatif	TTGACA	TATAAT
σ^{37}	utilisé dans la sporulation	AGGNTTT	GGNATTGNT
σ^{32}	utilisé dans la sporulation	AAATC	TANTGTTNTA
σ^{29}	synthétisé lors de la sporulation	TTNAAA	CATATT
σ^{28}	non utilisé dans la sporulation	CTNAAA	CCGATAT
gp ²⁸	expression intermédiaire SPO1	AGGAGA	TTTNTTT
gp ³³⁻³⁴	expression tardive SPO1	CGTTAGA	GATATT

La plupart de ces séquences consensus sont plutôt spéculatives ; seules les séquences pour l'enzyme de l'hôte (σ^{43}) et les deux enzymes du phage reposent sur un nombre suffisant de promoteurs cibles pour que le consensus soit fiable.

sigma pour diriger la reconnaissance de différents promoteurs ; ceci est rare chez *E. coli*, mais caractéristique de *B. subtilis*, bien que les holoenzymes des deux hôtes reconnaissent les promoteurs par les mêmes critères.

Parce que l'holoenzyme peut présenter (au moins) sept spécificités différentes chez *B. subtilis*, il semble peu probable que chaque facteur sigma puisse induire une conformation particulière dans l'enzyme « cœur » qui l'entraînerait à reconnaître un jeu distinct de promoteurs. Il est plus simple de raisonner en termes de facteurs sigma reconnaissant eux-mêmes directement les caractéristiques des promoteurs, bien que ceci pose le problème de savoir comment un petit polypeptide peut entrer en contact avec des sites couvrant plus de 20 pb d'ADN.

LES PROMOTEURS DE L'ADN POLYMERASE II SONT SITUÉS EN AMONT DU SITE D'INITIATION

Nous pouvons proposer plusieurs critères pour identifier les composants de la séquence d'un promoteur (ou de tout autre site de l'ADN) :

- Une séquence consensus doit être présente dans beaucoup d'exemples du site.
- Des mutations dans le site doivent bloquer le fonctionnement *in vitro* ou *in vivo*. (Il existe maintenant beaucoup de techniques pour introduire des mutations ponctuelles sur des paires de bases particulières, et en principe chaque site d'une séquence consensus ou toute autre région peut être mutée, et la séquence mutante testée *in vitro* ou *in vivo*.)
- Les protéines impliquées dans l'utilisation du site (tel que l'ARN polymérase ou des facteurs accessoires des promoteurs) doivent être liées à lui par la méthode d'empreinte.

Les tentatives pour définir les promoteurs de l'ARN polymérase des eucaryotes ont utilisé les renseignements obtenus précédemment avec les ARN polymérases bactériennes. Toutefois, les systèmes eucaryotes souffrent de deux limitations particulières :

- Pratiquement, aucune mutation sur le promoteur n'a été identifiée *in vivo* ; ainsi nous débutons sans aucune information antérieure sur la localisation du promoteur.
- Nous n'avons pas encore été en mesure de trouver la séquence d'ADN liée par une quelconque ARN polymérase. Recueillir des fragments protégés est rendu difficile par la complexité des préparations enzymatiques et le manque d'informations sur ce qui constitue

exactement la structure active de l'enzyme. Ce n'est bien sûr qu'une question de temps, pour que soit développé un système susceptible de nous permettre de purifier le site de liaison sur l'ADN à partir d'un complexe d'initiation*.

Un test moins direct, mais qui, actuellement, est en quelque sorte plus informatif, est de définir le promoteur par sa capacité à initier la transcription dans un système test convenable. Trois types de systèmes ont été utilisés :

- Le **système *in vitro*** reprend l'approche classique de la purification de tous les composants et les conditions de manipulation jusqu'à ce qu'une initiation fidèle soit obtenue. Une initiation « fidèle » est définie comme étant la production d'un ARN commençant au site correspondant à l'extrémité 5' de l'ARNm (ou de l'ARNr ou des précurseurs d'ARNt). Des systèmes pour chacune des trois ARN polymérases sont maintenant à des étapes différentes de purification ; à la fin chacun consistera en une préparation dans laquelle tous les composants auront été définis. Nous serons alors capable de comparer les activités *in vitro* des ARN polymérases de différents tissus ou espèces.
- Le **système de l'oocyte** suit les principes établis pour la traduction, et est basé sur l'injection d'un ADN matriciel convenable, cette fois dans le noyau d'un oocyte de *X. laevis*. L'ARN transcrit peut être purifié et analysé. La principale limitation de ce système est qu'il est limité aux conditions prévalant dans l'oocyte.
- Les **systèmes *in vivo*** n'impliquent pas nécessairement (comme le nom pourrait le laisser entendre) de caractériser la capacité d'une cellule de transcrire l'un de ses produits habituels. Ce qui est fait là, c'est de suivre la capacité d'une cellule à transcrire une matrice introduite par transfection (un procédé qui permet à un ADN exogène de pénétrer dans une cellule et d'être exprimé ; les événements impliqués sont détaillés dans le Chapitre 31). Le système est véritablement *in vivo*, à savoir que la transcription est accomplie par le même matériel que celui responsable de l'expression du propre génome de la cellule ; mais il diffère de la situation naturelle car la matrice peut consister en un gène qui ne serait pas habituellement transcrit dans la cellule de l'hôte.

L'approche pour caractériser le promoteur est la même dans les trois systèmes. Nous cherchons à manipuler la matrice *in vitro* avant qu'elle ne soit soumise au système de transcription. Cela est parfois appelé génétique substitutive.

* N. d. T. : Il est en fait possible que l'ARN polymérase II ne reconnaisse pas directement de séquence d'ADN spécifique mais plutôt des protéines fixées en des sites spécifiques de l'ADN.

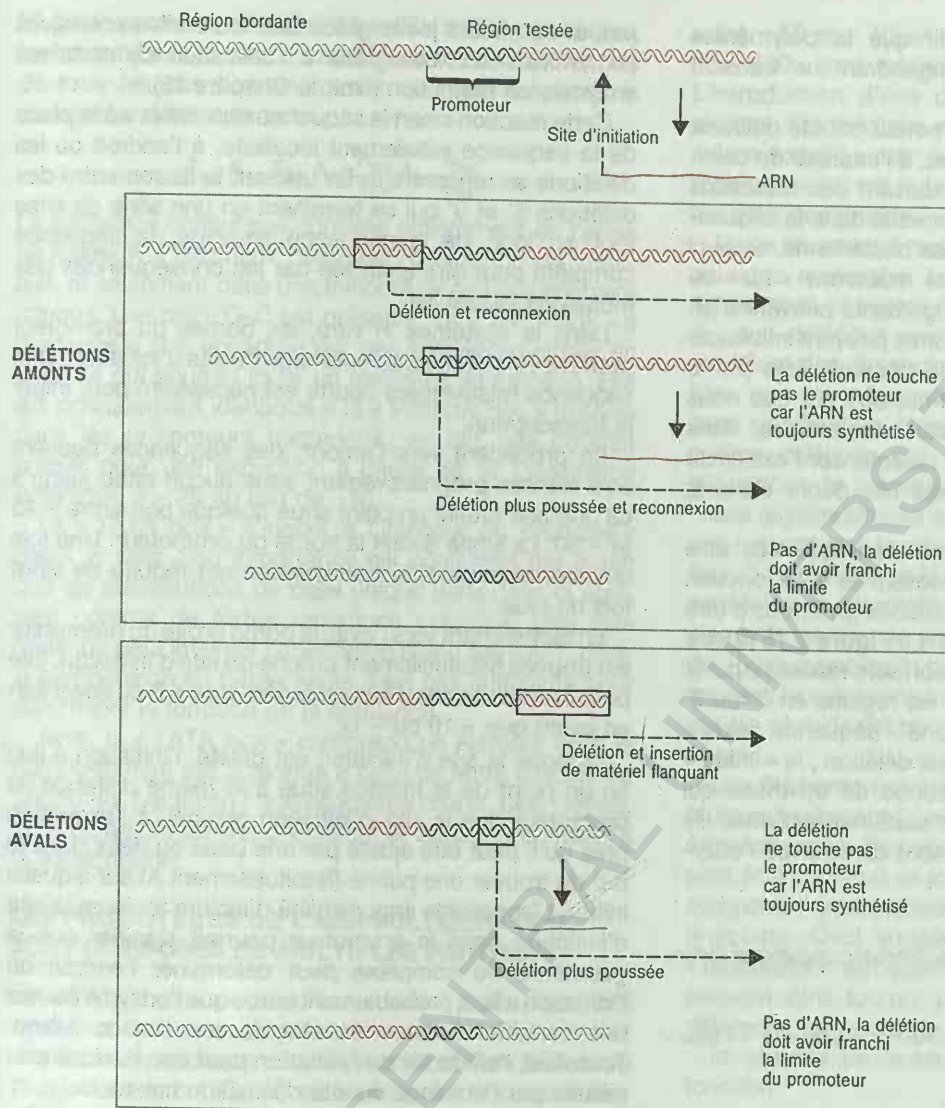


Figure 9.14
Les frontières du promoteur peuvent être déterminées par des délétions les encadrant.

Chaque délétion retire du matériel d'un côté de l'unité testée et s'étend plus avant dans l'unité que la précédente. Lorsqu'une délétion ne réussit pas à empêcher la synthèse de l'ARN mais que la suivante arrête la transcription, la limite du promoteur doit être située entre elles.

Lorsqu'un fragment d'ADN particulier peut être utilisé pour initier la transcription, il est considéré comme incluant un promoteur fonctionnel. Les limites de la séquence constituant ce promoteur peuvent alors être déterminées en réduisant la longueur du fragment à partir de chaque extrémité, jusqu'à ce qu'il cesse d'être actif, à partir d'un certain point. Le type de protocole est illustré dans la **Figure 9.14**. La limite amont peut être identifiée par délétion progressive de matériel à partir de cette extrémité jusqu'à ce que la fonction promoteur soit

perdue. Pour la limite aval, il est nécessaire de reconnecter le promoteur raccourci à la séquence devant être transcrite (autrement il n'y a plus de produit à tester).

Nous avons à prendre plusieurs précautions pour éviter des effets parasites. Pour être sûr que le promoteur est toujours dans le même contexte, la même séquence amont lui est toujours reconnectée. Parce que la terminaison peut ne pas se dérouler correctement dans les systèmes *in vitro*, la matrice peut être coupée à une certaine distance du promoteur (habituellement environ

500 pb en amont), pour garantir que la polymérase « décroche » au même point, engendrant un transcrit identifiable.

Une fois que les limites du promoteur ont été définies, l'importance de bases particulières, à l'intérieur de celui-ci, peut être déterminée en introduisant des mutations ponctuelles ou d'autres réarrangements dans la séquence. Comme avec l'ARN polymérase bactérienne, ceux-ci peuvent être caractérisés comme mutations « *up* » ou « *down* ». Certains de ces réarrangements peuvent n'affecter que le *taux* d'initiation ; d'autres peuvent influencer le *site* sur lequel l'initiation a lieu, observé alors par le changement du site d'initiation. Pour être sûr que nous avons bien affaire à des produits comparables, dans chaque cas, il est nécessaire de caractériser l'extrémité 5' de l'ARN, comme cela a déjà été décrit dans la Figure 9.7.

Le « **linker scanning** » est une technique très utile pour analyser les séquences nécessaires à la fonction promoteur ; elle permet à des mutations groupées d'être introduites à des sites particuliers. La Figure 9.15 illustre le protocole. Les mutants de délétion sont réalisés à partir des deux extrémités, en retirant les régions en 5' ou 3' par rapport au site considéré. Une « séquence linker » est ajoutée à l'extrémité de chaque délétion ; le « linker » consiste en un court oligonucléotide de synthèse qui inclut la séquence reconnue par certaines enzymes de restriction. Les deux fragments sont clivés avec l'enzy-

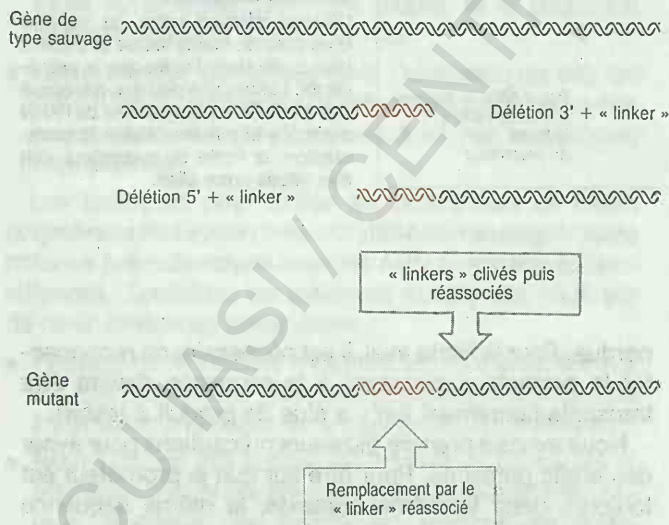


Figure 9.15

La technique de « linker-scanning » permet à une courte séquence de gène de type sauvage d'être remplacée par le « linker » à un endroit correspondant aux extrémités des délétions appariées.

me, et sont alors joints grâce aux extrémités cohésives (extrémités obtenues grâce à l'utilisation de certaines enzymes de restriction ; voir le Chapitre 15).

Cette réaction insert la séquence du « linker » à la place de la séquence initialement localisée, à l'endroit où les délétions se rencontrent. En utilisant la liaison entre des délétions 5' et 3' qui se terminent en une série de sites tout au long de la séquence sauvage, la séquence complète peut être analysée par les conséquences des mutations introduites.

Dans les systèmes *in vitro*, les bornes du promoteur peuvent être définies au voisinage du site d'initiation. Une séquence relativement courte est nécessaire pour initier la transcription.

En procédant vers l'amont, des séquences peuvent être retirées progressivement sans aucun effet, jusqu'à ce que soit atteint un point situé quelque part entre -45 et -30. La limite définit la borne du promoteur. Une fois la limite dépassée, la transcription est réduite de vingt fois ou plus.

En recherchant vers l'aval, la borne droite du promoteur est trouvée habituellement proche du site d'initiation. Elle peut être située aussi loin en arrière que +6 ou aussi loin en avant que -10 ou -12.

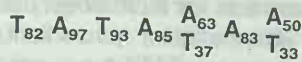
Lorsque le site d'initiation est délété, l'initiation a lieu en un point de la matrice situé à la même distance du promoteur que le site d'initiation originel, à l'exception près qu'il peut être ajusté par une base ou deux dans le but de trouver une purine (habituellement A) sur laquelle initier. L'apparente impossibilité d'inclure toujours le site d'initiation dans le promoteur pourrait signifier que la géométrie du complexe peut déterminer l'endroit où l'initiation a lieu, probablement parce que l'enzyme couvre une certaine distance en aval de son site de liaison. Toutefois, l'efficacité de l'initiation peut être quelque peu réduite par l'absence du site d'initiation habituel.

Comme avec les promoteurs bactériens, les homologies dans les régions proches du site d'initiation sont réduites à des séquences plutôt courtes. Des homologies potentiellement significatives ont été relevées dans trois régions : le site d'initiation, une région centrée à peu près -25, et une région plus éloignée en amont, centrée à peu près -75.

Au site d'initiation, il n'y a pas d'homologie extensive de séquence, mais la première base de l'ARNm a tendance à être une A, bordée de part et d'autre par des pyrimidines. (Cette description est également valable pour la séquence de démarrage CAT des promoteurs bactériens.)

La plupart des promoteurs possèdent une séquence appelée **TATA** ou « **Hogness box** », située entre 19 et

27 pb en amont du site d'initiation. Elle a été trouvée chez les mammifères, oiseaux, amphibiens et insectes. A partir de tous les cas connus (quelle que soit l'espèce), le consensus déterminé est



La séquence consensus consiste entièrement en paires de bases A-T (sur deux positions l'orientation est variable), et seulement dans une minorité de cas actuellement connus, une paire G-C est présente. La « TATA box » a tendance à être encadrée par des séquences riches en G-C, ce qui pourrait jouer un rôle dans sa fonction. Elle est pratiquement identique à la « Pribnow box » trouvée dans les promoteurs bactériens ; en fait, elle pourrait passer pour une exception à cause de sa localisation différente, à -25 au lieu de -10.

La « TATA box » s'inscrit à l'intérieur du promoteur défini *in vitro*. Son importance est confirmée par le fait que les substitutions de base unique dans celle-ci agissent comme de fortes mutations « down ». Une telle mutation peut inverser l'orientation d'une paire A-T, donc la composition en bases seule n'est pas suffisante pour déterminer la fonction de la séquence.

Ainsi, la « TATA box » comprend un élément dont le comportement est analogue à celui de notre modèle de promoteur bactérien : une séquence courte, bien définie, juste en amont du site d'initiation et qui est nécessaire à la transcription.

LES PROMOTEURS DE L'ARN POLYMÉRASE II SONT COMPOSÉS DE MULTIPLES PARTIES

La « TATA box » peut suffire *in vitro*, mais dans des systèmes *in vivo* ou *oocytaires*, nous constatons en plus une forte dépendance des séquences situées beaucoup plus loin en amont. Quelques courtes séquences consensus ont été identifiées, et des facteurs accessoires se liant à elles ont été trouvés.

Leur localisation exacte varie, mais ils sont situés à plus de 40 pb en amont de la « TATA box ». Lorsque cette région est délétée, l'initiation sur le site habituel est réduite à environ 2 % de son niveau initial. En revanche, si la « TATA box » est retirée, l'initiation continue d'avoir lieu, mais la localisation du site d'initiation diffère de celle du site initial qui est localisé très précisément.

En utilisant la technique de « linker scanning », trois éléments de séquences distincts ont été identifiés dans le gène de la thymidine kinase (TK) du virus herpès et dans le gène de la β -globine. La région proche de la « TATA box » est nécessaire à une initiation précise. Deux autres régions séparées, la *région intermédiaire* située

entre -50 et -70, et la *région distale* située entre -80 et -110, sont nécessaires à une initiation efficace. L'introduction d'une double mutation montre que les régions intermédiaire et distale sont concernées par la même fonction, car des mutations dans chacune des régions seules ont autant d'effet que des mutations dans les deux ensembles.

Cette structure à multiples parties signifie que nous ne pouvons rechercher la structure interne du promoteur simplement en créant des délétions à l'intérieur de celui-ci ; des délétions peuvent supprimer une fonction en modifiant des distances critiques entre des composants, bien que les séquences délétées soient elles-mêmes sans signification. Nous avons donc besoin de créer de petites délétions qui seront remplacées par d'autres séquences de même longueur.

Les régions situées entre les trois éléments de séquences ne sont pas impliquées dans la fonction du promoteur. Il apparaît que la distance entre les trois sites est modérément adaptable ; la séparation entre les deux éléments les plus en amont peut-être augmentée de plus de 15 pb avant qu'ils ne deviennent incapables de fonctionner ; leur distance à la « TATA box » peut être augmentée de plus de 30 pb avant que le promoteur n'en soit affecté.

Les éléments correspondants dans les deux promoteurs apparaissent jouer des rôles équivalents, et peuvent être échangés. Par exemple, un promoteur fonctionnel peut être construit en joignant le composant TK distal au composant intermédiaire et au composant TATA de la β -globine. Ceci suggère que le promoteur peut être « modulaire », en ce sens que ses composants individuels peuvent être fournis par des éléments de séquence différents, par exemple venant d'un autre promoteur.

In vivo, le promoteur semble avoir deux types de fonction :

- La fréquence d'initiation est fortement influencée par la *fonction amont* fournie par les éléments distal et intermédiaire. Ils ont probablement un effet majeur sur la liaison de l'ARN polymérase. Toutefois, la transcription résiduelle qui a lieu en leur absence est initiée sur le site d'initiation correcte.
- Le choix du site d'initiation dépend de la *fonction proximale*, portée par un élément voisin du site d'initiation qui encadre la « TATA box » dont la séquence est fortement conservée. Sa délétion fait que le site d'initiation devient variable, bien que la réduction globale de la transcription soit relativement faible. Le rôle de cette séquence pourrait être de positionner l'ARN polymérase de manière à ce qu'elle initie sur le site approprié.

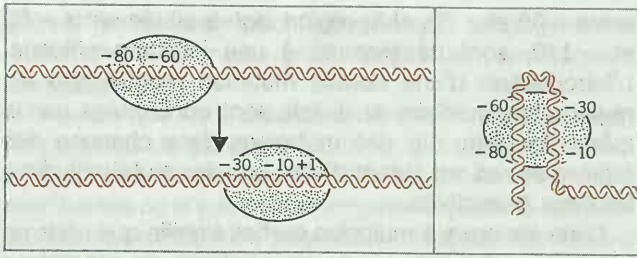


Figure 9.16

Un des modèles permettant d'accorder la taille du promoteur avec celle de l'ARN polymérase consiste à supposer que l'enzyme se déplace. Une autre hypothèse est que l'ADN est organisé de manière compacte.

Comment un promoteur peut-il être constitué d'éléments séparés, s'étendant sur une longueur d'ADN plus grande que celle que l'ARN polymérase peut couvrir ? Deux modèles sont illustrés dans la Figure 9.16.

Il est possible que l'ARN polymérase entre en contact d'abord avec le site le plus éloigné en amont puis se déplace vers le site le plus proche du site d'initiation. Une autre possibilité, si l'on se rappelle que, *in vivo* l'ADN n'est pas allongé de manière linéaire, est que son organisation compacte pourrait amener les sites séparés dans la molécule duplexe à se juxtaposer (voir le Chapitre 26). Le site de liaison de l'ARN polymérase pourrait consister en séquences d'ADN non contiguës, mais maintenues ensembles par des protéines liant l'ADN. Ce modèle implique de plus que l'espace entre les composants du promoteur (plutôt que la séquence exacte) est important.

Dans chaque type de modèle, nous constatons que la région la plus en amont pourrait être la plus importante pour lier initialement l'ARN polymérase, mais la région la

plus proche du site d'initiation pourrait être requise pour maintenir l'enzyme dans une configuration lui permettant de reconnaître le site exact d'initiation. Ainsi, lorsque la séquence la plus en amont est absente, la capacité de l'ARN polymérase à se lier à l'ADN est très réduite ; mais ces molécules enzymatiques qui se lient peuvent initier très précisément car la « TATA box » est présente. D'un autre côté, lorsque la « TATA box » est délétée, l'ARN polymérase reste capable de se lier efficacement à la séquence amont, mais ses contacts dans la région du site d'initiation sont beaucoup moins précis, ce qui permet à l'initiation d'avoir lieu en plus d'un point.

Comment pouvons-nous expliquer la contradiction entre le besoin *in vivo* de la séquence amont - 50 et son apparent manque d'importance *in vitro* ? La différence d'efficacité de transcription *in vivo* et *in vitro* pourrait être principalement responsable de cet effet. Le système *in vitro* est relativement inefficace ; moins de 1 % de la matrice est réellement transcrit. Nous ignorons exactement quelle proportion de la matrice est utilisée *in vivo* ou dans l'oocyte, mais elle pourrait très bien être beaucoup plus élevée. Le niveau *in vitro* pourrait même correspondre à l'expression résiduelle constatée *in vivo* en l'absence de séquences amont à - 60 !

En tout cas, la structure matricielle la plus simple *in vitro*, telle qu'une molécule d'ADN non organisée dans une structure protéique courante, peut signifier que sa reconnaissance par l'ARN polymérase est différente, et certainement moins efficace, peut-être parce qu'elle n'utilise seulement que quelques propriétés du promoteur*.

Quelles séquences actives composent les modules du promoteur ? La fonction de chaque module est déterminée par une séquence consensus plutôt courte, bien que la séquence encadrant ce « cœur » puisse influencer son

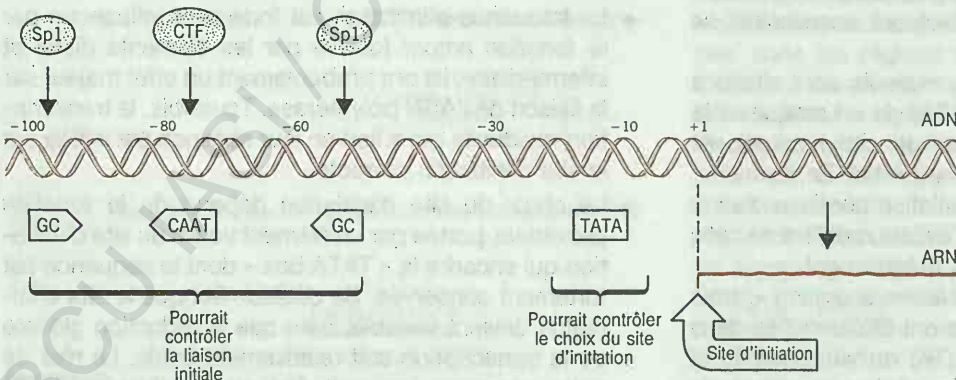


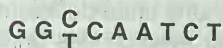
Figure 9.17

Un promoteur pour l'ARN polymérase II comporte des composants de séquences séparées, possédant des rôles différents. Les séquences entre les composants ne sont pas importantes. L'orientation des séquences conservées est indiquée par la direction des flèches. Les facteurs de transcription se lient aux motifs de séquences conservées.

* N. d. T. : Depuis 1986, l'utilisation de systèmes plus efficaces de transcription *in vitro* a permis d'étudier l'influence d'éléments d'ADN situés relativement loin en amont du site d'initiation, par exemple à plus de 100 bases de celui-ci.

efficacité. Les séquences consensus sont toutes décrites en terme de « boîtes », tel que la « boîte TATA » (« TATA box »).

La « **CAAT box** » est trouvée à la fois dans les promoteurs TK et β -globine et est conservée dans plusieurs promoteurs connus (mais pas dans tous). Elle possède le consensus suivant :



Elle est présente telle quelle dans la région intermédiaire de la β -globine, et dans une orientation inversée dans la région distale de TK.

La « **GC box** » constitue un autre élément conservé, contenant la séquence GGGCGG. De multiples copies sont souvent présentes dans l'une ou l'autre orientation, au niveau des promoteurs.

L'organisation de ces éléments dans le promoteur de la thymidine kinase est résumée dans la **Figure 9.17**. En procédant vers l'amont, à partir du site d'initiation, on constate que la « TATA box » se trouve dans la région proximale, la région intermédiaire contient une « GC box », et la région éloignée, une « CAAT box » ainsi qu'une seconde « GC box ». Les deux « GC boxes » ont des orientations opposées, et la « CAAT box » est orientée à l'inverse de l'accoutumé. L'une des énigmes touchant à l'organisation du promoteur est que celui-ci porte une information directionnelle (la transcription ne se déroule que vers l'aval), mais les « GC box » et « CAAT box » semblent capables de fonctionner dans les deux orientations.

Aucun de ces éléments n'est présent dans tous les promoteurs. Quelques promoteurs ne possèdent pas de « TATA box », et dans ceux-ci, l'initiation ne débute pas à un site unique, mais à l'un ou l'autre d'un ensemble de

sites d'initiation. D'autres n'ont pas de « CAAT box » et/ou n'ont pas de « GC boxes ». Des exemples de composants de certains promoteurs sont résumés dans la **Figure 9.18**.

Des promoteurs pour l'ARN polymérase II dans la levure ont une « TATA box » identique au promoteur d'eucaryote supérieur que nous venons de décrire, mais ils en diffèrent car ils tolèrent une large variation de sa localisation, de 40-90 pb en amont du site d'initiation. La « TATA box » est essentielle à la transcription, mais des séquences au, ou à proximité du site d'initiation sont nécessaires pour fixer le point précis de l'initiation.

LES FACTEURS DE TRANSCRIPTION RECONNAISSENT DES SÉQUENCES CONSENSUS PARTICULIÈRES

Le critère permettant de dire qu'une protéine fait partie de l'appareil transcriptionnel est qu'elle est nécessaire pour que la transcription soit initiée dans un système *in vitro*. Les protéines de l'appareil transcriptionnel peuvent être divisées en trois groupes :

- Les sous-unités de l'ARN polymérase sont nécessaires à certaines ou à toutes les étapes de la transcription, mais elles ne sont pas spécifiques de promoteurs particuliers.
- Les facteurs transcriptionnels peuvent se lier à l'ARN polymérase pendant ou après qu'elle ait formé un complexe d'initiation, bien qu'ils ne fassent pas partie de l'enzyme libre. Ces facteurs sont vraisemblablement nécessaires à l'initiation de la transcription au niveau de tous les promoteurs ou (par exemple) à la terminaison. Ceci nous ramène immédiatement au problème

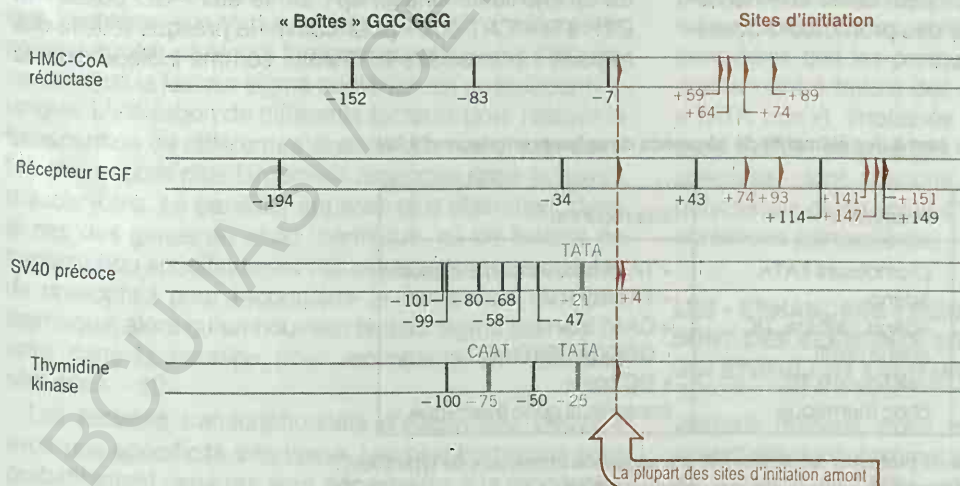


Figure 9.18
Les promoteurs contiennent différentes combinaisons de « TATA boxes », « CAAT boxes » et « GC boxes ».

suivant : quels polypeptides sont des sous-unités de l'ARN polymérase et quels autres sont des facteurs accessoires.

- Les facteurs de transcription peuvent se lier à des séquences spécifiques dans le promoteur cible. Si ces séquences sont présentes dans tous les promoteurs, ces facteurs font partie de l'appareil transcriptionnel général. Si certaines séquences sont présentes seulement dans certaines classes de promoteurs, des facteurs les reconnaissant peuvent être nécessaires pour initier spécifiquement sur ces promoteurs.

La plupart des efforts réalisés pour caractériser les facteurs transcriptionnels ont porté sur la dernière catégorie, et nous connaissons maintenant certains facteurs qui sont nécessaires à la transcription spécifique à partir de promoteurs contenant des séquences consensus particulières. Ils sont résumés dans le **Tableau 9.6**.

La caractérisation des sites de liaisons de ces facteurs nous fournit un critère supplémentaire pour identifier les composants du promoteur. Nous nous attendons à ce qu'un facteur se lie à une région de l'ADN incluant une séquence identifiée comme essentielle pour la transcription *in vitro* par les analyses de mutation.

Deux facteurs au moins sont nécessaires à l'ARN polymérase de *D. melanogaster* pour initier la transcription *in vitro*. L'un est le facteur B, qui se lie à une région contenant la « TATA box ». Le facteur humain TFIID se lie à une région similaire. *Quelques traits caractéristiques des promoteurs peuvent donc permettre une interaction directe avec des facteurs (tels que ceux séparés par chromatographie) plutôt qu'avec l'ARN polymérase II elle-même.*

De façon similaire, le CTF (facteur se liant à CAAT) se lie à la région du promoteur tardif majeur d'adénovirus homologue à la « CAAT box » ; il peut donc être nécessaire pour la transcription à partir des promoteurs possédant cette séquence consensus*.

USF (également appelé MLTF) est un autre facteur se liant dans la région amont, il reconnaît une séquence située vers - 55 dans le promoteur tardif de l'adénovirus.

La présence de la « GC box » est associée à la capacité de lier le facteur de transcription sp1. Dans le promoteur SV40, les multiples « boîtes » entre - 70 et - 110 sont toutes liées, ainsi la région dans son ensemble est protégée par sp1. Une unité de liaison sp1 entre probablement en contact avec un des brins de l'ADN, sur un site de liaison d'environ 20 pb qui inclut au moins une « GC box ». Nous ne connaissons pas encore la forme active de la protéine, qui possède une sous-unité monomérique de 100 000 daltons**.

Bien que la « GC box » soit essentielle pour la liaison, différentes séquences contenant des « GC boxes » sont liées avec des affinités différentes ; ainsi la séquence flanquante peut influencer la reconnaissance par sp1. Nous ignorons quelle fonction la liaison sp1-« GC box » joue dans la transcription, et comment et pourquoi les « GC boxes » peuvent fonctionner à des distances variables du site d'initiation.

Le facteur sp1 pourrait rendre l'ARN polymérase II capable de reconnaître une certaine classe de promoteurs. Certaines hypothèses proposent que les « GC boxes » pourraient notamment identifier des gènes exprimés de manière constitutive.

Les promoteurs pourraient être classés en différents groupes caractérisés par la présence de séquences consensus particulières liant ces facteurs. Des facteurs reconnaissant ces séquences pourraient en effet être responsables de la reconnaissance coordonnée de classes de promoteurs.

La reconnaissance par des facteurs multiples peut être nécessaire pour initier la transcription. Dans le promoteur de la thymidine kinase, sp1 se lie aux « GC boxes » et CTF à la « CAT box » pour couvrir la presque totalité des régions intermédiaire et distale, comme indiqué dans la

Tableau 9.6

Certains facteurs transcriptionnels se lient à des éléments de séquence dans les promoteurs d'ARN polymérase II.

Facteur	Espèces	Promoteur	Site reconnu
B	Drosophile	promoteurs TATA	« TATA box »/site d'initiation
TFIID	homme	adéno	« TATA box »
CTF	homme	adéno, globine, TK	« CAAT box »
USF/MLTF	homme	adéno tardif	GGCCACGTGACC
sp1	homme	voir figure 9.18	« GC box »
HSTF	Drosophile	choc thermique	consensus choc thermique

* N. d. T. : Il existe toute une famille de protéines reconnaissant une séquence consensus de type CAAT.

** N. d. T. : La protéine sp1 a été totalement purifiée et son gène récemment cloné.

Figure 9.17. Dans le promoteur tardif de l'adénovirus, TFIID se lie à la « TATA box » et USF à la région voisine.

Lorsque plusieurs facteurs se lient, leurs interactions avec le promoteur peuvent être coopératives, de telle manière que la liaison d'un facteur peut aider à celle d'un autre facteur. Nous pouvons concevoir leurs interactions sur le promoteur comme élaborant une structure qui, avec l'ARN polymérase, constitue un complexe d'initiation.

Ces facteurs sont nécessaires à la transcription générale, c'est-à-dire que nous ne les impliquons pas dans une capacité de régulation. D'autres facteurs peuvent être utilisés pour réguler la transcription de groupes de gènes particuliers. Un bon exemple est donné par les gènes du choc thermique.

La réponse au choc thermique est commune à une large variété de procaryotes et d'eucaryotes et peut impliquer de multiples contrôles de l'expression des gènes : une augmentation de température éteint la transcription de certains gènes, active celle des **gènes du choc thermique**, et peut également provoquer des changements dans la traduction des ARNm. Les gènes du choc thermique d'eucaryotes possèdent une séquence consensus d'environ 15 pb en amont du site d'initiation. Le facteur HSTF est actif seulement dans les cellules ayant subi un choc thermique ; il se lie à un site incluant la séquence consensus du choc thermique. L'activation de ce facteur, par conséquent, fournit un moyen d'initier la transcription d'un groupe spécifique d'environ 20 gènes qui contiennent la séquence cible appropriée sur leur promoteur.

Nous ne comprenons pas le mécanisme de l'initiation avec suffisamment de détails pour savoir quel rôle jouent les facteurs, mais en terme d'effet global, nous pouvons considérer la séparation entre l'ARN polymérase II et les facteurs comme étant analogue à celle existant entre l'enzyme « cœur » bactérien et le facteur sigma. Le fait que l'initiation sur tout promoteur donné d'eucaryotes puisse requérir plusieurs facteurs constitue une différence, puisque le facteur sigma bactérien est un polypeptide unique. L'utilisation de différents facteurs pour réaliser la transcription de différentes classes de promoteurs semble être encore plus largement répandue dans le noyau d'eucaryotes. Le parallèle apparaît plus clairement dans le cas des gènes du choc thermique, où un facteur de transcription spécifique est nécessaire dans des cellules de drosophile pour reconnaître le consensus du choc thermique, alors qu'un nouveau facteur sigma est nécessaire dans la bactérie pour reconnaître une nouvelle séquence - 10.

Les facteurs transcriptionnels d'eucaryotes peuvent avoir une spécificité très vaste. Les plus communs sont probablement ceux qui sont nécessaires à la reconnais-

sance de tous les promoteurs contenant des « TATA boxes ». Les facteurs les moins communs peuvent être ceux utilisés pour activer la transcription de petits groupes de gènes répondant à certains signaux communs. Ainsi la diversité des facteurs peut s'étendre des composants de l'appareil transcriptionnel relativement communs jusqu'aux protéines de régulation relativement rares.

Cette idée peut être étendue à un principe général. *Un gène peut être régulé par une séquence sur le promoteur qui est reconnue par une protéine spécifique. La protéine fonctionne comme un facteur de transcription nécessaire à l'ARN polymérase pour initier la transcription. Elle est disponible seulement dans certaines conditions, lorsque le gène va être exprimé ; son absence assure que le promoteur ne peut pas être utilisé.* Pourquoi le promoteur est-il dépendant de la reconnaissance de cette séquence ? Peut-être manque-t-il d'un ensemble complet de modules nécessaires à la reconnaissance par l'ARN polymérase II ; le facteur se liant à la séquence de régulation compenserait alors cette déficience.

Nous pouvons mener ce principe plus loin. *Lorsqu'un promoteur est régulé de plus d'une manière différente, chaque événement régulationnel peut dépendre de la liaison de sa propre protéine à une séquence particulière.* Par exemple, l'un des gènes du choc thermique de la drosophile est exprimé dans deux circonstances : dans tous les tissus à température élevée et dans les ovaires à température normale. Une analyse délétionnelle montre que des séquences de - 49 à - 85 sont nécessaires pour l'initiation durant le choc thermique, alors que des séquences en amont de - 341 sont nécessaires pour l'expression ovarienne.

Ces éléments de régulation sont habituellement plutôt courts. Par exemple, une séquence d'environ 12 pb permet au promoteur de la métallothionéine d'être activé en présence d'ions métalliques. Nous pouvons alors considérer que les promoteurs des ARN polymérases II contiennent à la fois des modules de bases (tels que la « TATA box »), impliqués dans une interaction générale avec l'ARN polymérase, et une série de séquences spéciales, dont chacune d'entre elles permet à l'ARN polymérase de se lier et d'initier la transcription dans des conditions particulières.

LES « ENHANCERS » (SÉQUENCES ACTIVATRICES) SONT DES ÉLÉMENTS BIDIRECTIONNELS QUI STIMULENT L'INITIATION

Jusqu'à présent, nous avons considéré le promoteur essentiellement comme une région isolée, responsable de la liaison de l'ARN polymérase. Mais les promoteurs

d'eucaryotes ne fonctionnent pas nécessairement seuls. Dans certains cas au moins, l'activité d'un promoteur est énormément augmentée par la présence d'une autre séquence, connue sous le nom de « **enhancer** ».

Un « **enhancer** » diffère du promoteur lui-même par deux caractéristiques : sa position par rapport au promoteur n'a pas besoin d'être fixée, mais elle peut varier substantiellement ; et il peut fonctionner dans chaque orientation. Un « **enhancer** » n'est pas limité à l'assistance d'un promoteur particulier, mais il peut stimuler n'importe quel promoteur placé dans son voisinage.

Le « **enhancer** » de SV40 est localisé dans une région du génome contenant deux séquences identiques de 72 pb chacune, répétées en tandem à environ 200 pb en amont du site d'initiation d'une unité de transcription. Ces **72 pb répétées** sont situées dans une région possédant une structure nucléoprotéique inhabituelle : elle est apparemment plus exposée que d'habitude (voir le Chapitre 27). La carte délétionnelle montre que chacune de ces séquences répétées est capable d'assurer une transcription normale ; mais le retrait de ces deux séquences réduit grandement la transcription *in vivo*.

Avec ce type de critères, nous pouvons postuler que la région répétée constitue un composant amont du promoteur. Mais des expériences de reconstruction dans lesquelles la séquence de 72 pb est retirée de l'ADN puis remplacée ailleurs, montrent qu'une transcription normale peut avoir lieu aussi longtemps que cette séquence est présente *quelque part* sur la molécule d'ADN. En fait, si un gène de la β -globine est placé dans une molécule d'ADN contenant la séquence répétée de 72 pb, sa transcription est augmentée *in vivo* de plus de 200 fois, même lorsque la séquence 72 pb est située aussi loin que 1 400 pb en amont ou 3 300 pb en aval du site d'initiation. Et ceci constitue simplement les limites testées jusqu'à présent ; il nous reste à découvrir à quelle distance la séquence – 72 pb cesse d'être fonctionnelle. La séquence peut être inversée et remplacée et être toujours fonctionnelle.

La morale de ces résultats est claire. Nous devons nous montrer prudents dans la définition des composants du promoteur. Il ne suffit pas seulement de montrer qu'une délétion sur une séquence particulière réduit la transcription ; nous devons également étudier l'importance de la *localisation* de la séquence délétée.

Or donc, qu'est-ce qu'un promoteur ? *Si nous utilisons une définition fonctionnelle telle que : il est constitué d'une ou de plusieurs séquences d'ADN qui doivent être situées de manière (relativement) fixe par rapport au site d'initiation, la « TATA box » et d'autres éléments amont sont inclus, mais le « enhancer » est exclu.*

Plusieurs génomes viraux incluent des « **enhancers** ».

Les rétrovirus présentent un intérêt particulier ; ce sont des virus dont l'insertion dans le génome de l'hôte peut activer des gènes à proximité, peut-être par l'intermédiaire de la présence d'un « **enhancer** ». Certains « **enhancers** » viraux montrent une spécificité dans leur fonction pour un type cellulaire. La capacité qu'a un « **enhancer** » de fonctionner dans une cellule particulière peut être en partie responsable de la détermination de la spécificité d'hôtes du virus. Dans le cas du polyome, une mutation dans le « **enhancer** » élargit la gamme des cellules que le virus peut infecter.

Des « **enhancers** » cellulaires ont été découverts dans le génome, sous la forme d'éléments stimulant l'utilisation d'un promoteur proche dans un tissu donné. De tels « **enhancers** » peuvent fournir une contribution au réseau de régulation par lequel l'expression des gènes est contrôlée. Un cas particulier est représenté par les gènes d'immunoglobulines, qui comportent des « **enhancers** » à l'intérieur de l'unité de transcription. Ainsi le « **enhancer** » est en aval du promoteur qui est stimulé. Les « **enhancers** » d'immunoglobulines apparaissent être actifs seulement dans les lymphocytes B dans lesquels les gènes d'immunoglobulines sont exprimés (voir le Chapitre 32).

Il nous reste à délimiter les bornes des « **enhancers** ». Des substitutions de bases créent des mutations « **down** » sur une distance d'environ 100 pb, recouvrant la séquence répétée de 72 pb de SV40. Des séquences consensus courtes peuvent être déduites à partir de la séquence des « **enhancers** » connus, mais les arguments en faveur de leur rôle fonctionnel sont rarement évidents.

La meilleure preuve de l'implication de séquences particulières est fournie par des expériences d'empreinte sur des « **enhancers** » *in vivo*. Lorsque l'agent de méthylation diméthylsulfate (DMS) est appliqué *in vivo*, il réagit avec l'ADN de la même manière qu'*in vitro*. Utilisé *in vivo*, il identifie une séquence d'ADN liée à une protéine grâce à la protection ou à l'augmentation de la méthylation de résidus G particuliers par rapport à ce que l'on observe *in vitro* sur de l'ADN libre. Naturellement, ceci ne nous dit pas *quelle protéine* se lie à l'ADN *in vivo*.

Un « **enhancer** » d'immunoglobuline montre un ensemble de protection et d'hyperméthylation au niveau de variantes d'une séquence consensus octamérique, CAGGTGGC. Quatre octamères sont présents dans la région « **enhancer** ». La réaction n'a lieu que dans les lymphocytes B, ce qui suggère que le « **enhancer** » peut être activé par une protéine spécifiquement synthétisée dans ces cellules. Une séquence en rapport avec ce consensus est trouvée dans plusieurs « **enhancers** ».

Des « **enhancers** » peuvent avoir des structures modulaires. Le « **enhancer** » de SV40 peut être séparé en deux

moitiés, aucune d'entre elles ne fonctionnant par elle-même, mais les deux étant actives même séparées par une certaine distance. Une analyse mutationnelle identifie trois domaines : une mutation de l'un d'entre eux inactive le « enhancer », mais elle peut être annulée par la duplication de l'un ou l'autre des deux domaines restants. Bien que les domaines soient de séquences différentes, ils apparaissent jouer des rôles similaires, puisqu'un « enhancer » actif peut être créé par la combinaison d'un nombre suffisant de domaines de type sauvage, indépendamment de leurs types.

Comment un « enhancer » peut-il stimuler l'initiation sur un promoteur qui peut être situé de n'importe quel côté de lui et apparemment à n'importe quelle distance ? Il est possible que cela ait à voir avec la structure, la localisation ou la liaison de l'enzyme :

- Un « enhancer » pourrait modifier la structure globale de la matrice, par exemple, en influençant l'organisation ADN-protéine de la chromatine, ou en changeant la densité de superenroulement. En défaveur de cette hypothèse est le fait qu'un « enhancer » continue de fonctionner dans des conditions où le superenroulement ne peut pas être propagé le long de la double hélice.

Tous les « enhancers » caractérisés jusqu'à maintenant incluent une région alternant les résidus pyrimidine-purine, le type même de séquence susceptible de former de l'ADN-Z (voir le Chapitre 3). Il est possible que les « enhancers » fonctionnent en formant un court segment d'ADN-Z, bien que nous restions dans l'expectative quant à l'effet de l'ADN-Z sur une transcription locale. Toutefois, un effet structural de cette nature expliquerait pourquoi le « enhancer » peut affecter un promoteur de chaque côté.

- Si un « enhancer » fournit un site d'attachement, il pourrait être responsable de la localisation de la matrice non à une place particulière à l'intérieur de la cellule, par exemple en l'attachant à la matrice nucléaire.
- Un « enhancer » pourrait fournir un « site d'entrée » bidirectionnel, un point sur lequel l'ARN polymérase (ou une autre protéine essentielle) s'associerait à la chromatine. La polymérase devrait alors se déplacer vers le promoteur.

Des éléments quelque peu similaires aux « enhancers », appelés séquences activatrices amont (UAS, upstream activator sequences), sont trouvés dans la levure. Ils peuvent fonctionner dans chaque orientation,

à des distances variables en amont du promoteur, mais ne peuvent fonctionner lorsqu'ils sont situés en aval. Ils ont un rôle de régulation : dans le système *gal*, le UAS est lié par une protéine régulatrice qui active les gènes.

Certaines expériences indiquent que si un site lié par une autre protéine, ou une séquence entraînant la fin de la transcription, est introduite entre le UAS et le promoteur du gène, l'effet d'activation est aboli. Ceci serait cohérent avec un modèle de cheminement dans lequel certaines protéines, peut-être l'ARN polymérase elle-même, se déplaceraient physiquement vers le promoteur, ce déplacement étant sensible à un blocage physique par une autre protéine ou à des signaux de régulation commandant une terminaison du mouvement.

La généralité du phénomène « enhancer » n'est pas claire. Nous ignorons quelle proportion de promoteurs cellulaires dépend habituellement d'un « enhancer » pour atteindre leur niveau d'expression usuel. Nous ne savons pas non plus si un « enhancer » fournit souvent une cible pour la régulation. Certains « enhancers » sont activés seulement dans les tissus dans lesquels leurs gènes fonctionnent, mais d'autres pourraient être actifs dans toutes les cellules.

Dans un cas particulièrement remarquable, un « enhancer » est responsable d'une réponse hormonale. La transcription de l'ADN du virus responsable de la tumeur mammaire de la souris est stimulée par les hormones stéroïdes. L'élément responsable de la réponse hormonale est situé à environ 100 pb en amont du site d'initia-

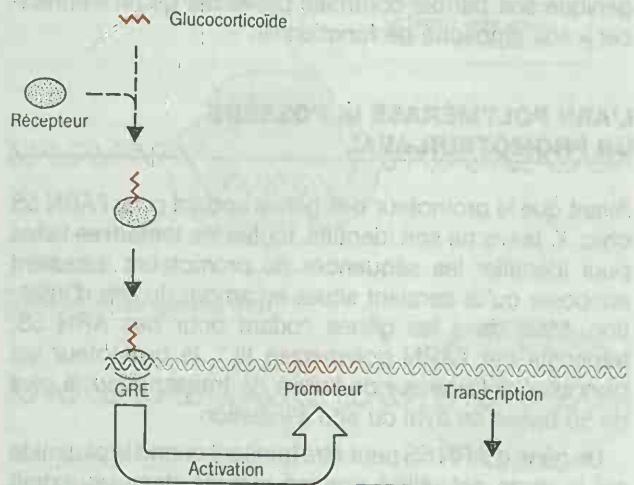


Figure 9.19

Les glucocorticoïdes régulent la transcription de gènes en faisant se lier leurs récepteurs à un « enhancer » dont l'action est nécessaire à la fonction du promoteur.

* N. d. T. : Des expériences plus récentes sur le mécanisme d'activation des gènes du système *gal* par l'intermédiaire des UAS favorisent le modèle d'un contact direct entre la ou les protéines liées à l'UAS et l'ARN polymérase liée au niveau du promoteur. Ce contact direct se ferait grâce à la formation de boucles d'ADN.

tion, liant un complexe constitué de l'hormone, et de son récepteur protéique ; il peut stimuler la fonction d'autres gènes lorsqu'il est placé dans l'une ou l'autre orientation à des distances variables de leur promoteur. En bref, il se comporte comme un « enhancer » associé à une séquence de fixation hormonale.

L'activation du « enhancer » peut être le mécanisme général par lequel les stéroïdes régulent un ensemble de gènes cibles. La **Figure 9.19** illustre les processus impliquant les glucocorticoïdes et pour lesquels nous possédons le plus d'informations. Les glucocorticoïdes pénètrent dans la cellule et se lient à leurs récepteurs. La liaison active le récepteur qui reconnaît alors une séquence consensus présente dans des « enhancers » situés près des gènes répondant aux glucocorticoïdes. Lorsque le complexe glucocorticoïde-récepteur se lie au « enhancer », le promoteur voisin est activé, et la transcription est initiée à cet endroit.

Bien qu'une séquence consensus soit commune aux « enhancers » régulés par les glucocorticoïdes, la longueur de la séquence liée par le récepteur et sa localisation par rapport au promoteur sont variables. La séquence consensus est essentielle pour la liaison, mais les séquences environnantes sont également nécessaires. L'élément de la réponse aux glucocorticoïdes (GRE, glucocorticoid response element) peut se trouver à plusieurs kilobases en amont, ou en aval du promoteur.

Dans certains cas, l'activité d'un « enhancer » est inhibée spécifiquement ; il est donc possible que l'activité génique soit parfois contrôlée par le fait qu'un « enhancer » soit empêché de fonctionner.

L'ARN POLYMERASE III POSSÈDE UN PROMOTEUR AVAL

Avant que le promoteur des gènes codant pour l'ARN 5S chez *X. laevis* ne soit identifié, toutes les tentatives faites pour identifier les séquences de promoteurs laissaient supposer qu'ils seraient situés en amont du site d'initiation. Mais dans les gènes codant pour des ARN 5S, transcrits par l'ARN polymérase III*, le promoteur est bien situé à l'intérieur de l'unité de transcription, à plus de 50 bases en aval du site d'initiation.

Un gène d'ARN 5S peut être transcrit quand le plasmide qui le porte est utilisé comme matrice dans un extrait nucléaire obtenu à partir d'oocyte de *X. laevis*, le tissu dans lequel le gène est habituellement exprimé. Le promoteur a été localisé en utilisant des plasmides dans

lesquels des délétions s'étendent dans le gène, dans chaque direction. Le produit ARN 5S continue d'être synthétisé lorsque la séquence entière en amont du gène est retirée.

Lorsque les délétions se prolongent dans le gène, un produit très similaire en taille à l'ARN 5S habituellement obtenu continue d'être synthétisé tant que la délétion s'arrête avant à peu près la base + 55. La première partie de l'ARN produit représente l'ADN du plasmide ; la seconde partie représente tout segment restant de la séquence habituelle de l'ARN 5S. Mais lorsque la délétion dépasse + 55, la transcription n'a pas lieu. Ainsi, le promoteur est situé *en aval d'une position + 55*, mais fait que l'ARN polymérase III initie la transcription en un site plus ou moins fixe à distance. Le site d'initiation sauvage est unique ; dans les délétions qui l'ont perdu, la transcription est initiée sur la base purique la plus proche de la position à + 55 pb en amont du promoteur.

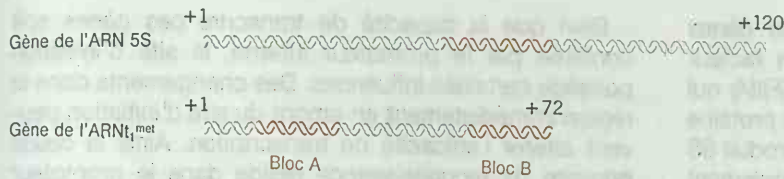
Lorsque des délétions s'étendent dans le gène à partir de son extrémité distale, la transcription n'est pas affectée tant que les premières 80 pb restent intactes. Dès que la délétion empiète sur cette région, la transcription cesse. Ceci localise la borne aval du promoteur à une position située à à peu près + 80.

Ainsi, le promoteur pour la transcription de l'ARN 5S se situe entre les positions + 55 et + 80, à l'intérieur du gène. Un fragment contenant cette région peut garantir l'initiation de n'importe quel ADN dans lequel il est placé, celle-ci ayant lieu à une position d'environ 55 pb en amont. Comment l'ARN polymérase initie-t-elle la transcription en amont de son promoteur ? L'explication la plus vraisemblable est que l'enzyme se lie au promoteur, mais est assez large pour entrer en contact avec des régions situées à 55 pb.

Comme avec l'ARN polymérase II, la géométrie de la liaison sur le promoteur doit conditionner la position du site d'initiation, à la restriction près que les pyrimidines ne peuvent pas être utilisées pour l'initiation. La différence entre ces enzymes est que l'ARN polymérase II atteint le site d'initiation à partir de son promoteur, loin en avant, alors que l'ARN polymérase III l'atteint en arrière.

Les bornes de certains promoteurs d'ARN polymérase III sont résumées dans la **Figure 9.20**. Elles s'inscrivent toujours dans les limites de l'unité de transcription. Dans les gènes d'ARNt, le promoteur se présente en deux parties séparées, toutes deux situées à l'intérieur du gène. La carte délétionnelle montre que les deux séquences, celle du *bloc A*, située entre + 8 et + 30, et celle du *bloc B*, située entre + 51 et + 72, doivent être présentes.

* N. D. T. : Depuis la fin 1987, quelques exemples ont été décrits de promoteurs pour l'ARN polymérase III situés immédiatement en amont de l'unité de transcription.

**Figure 9.20**

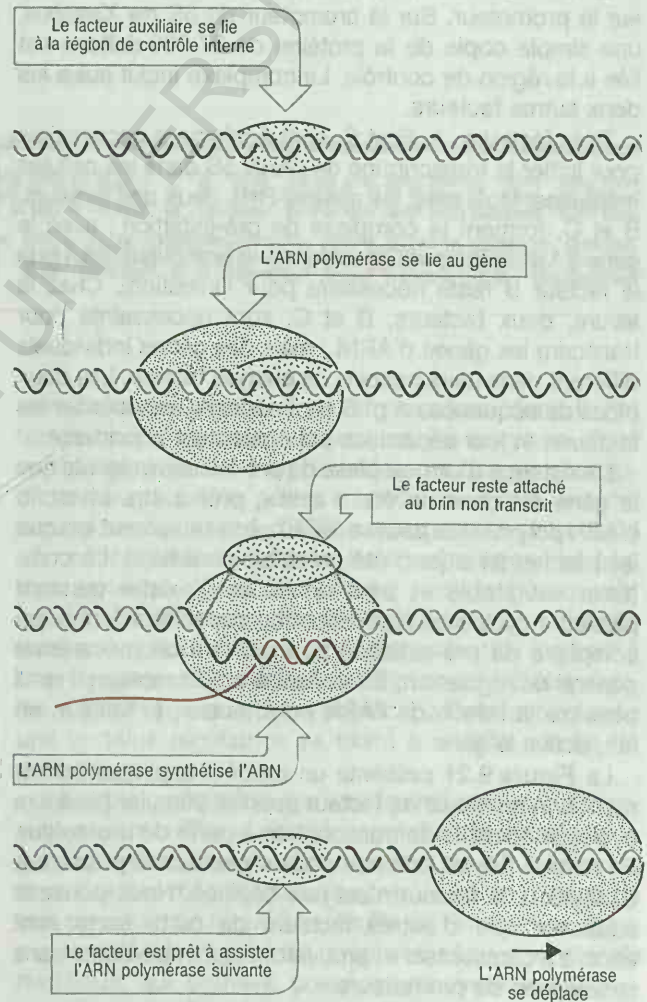
Les gènes reconnus par l'ARN polymérase III contiennent tous des promoteurs internes (régions en rouge). La région de contrôle interne peut être un bloc unique (comme dans les gènes de 5S) ou deux blocs séparés (comme dans les gènes de l'ARNt). Le nombre indique la première et la dernière base du gène.

Des changements dans la séquence entre les blocs sont sans effet. Toute délétion réduisant leur distance de séparation empêche l'initiation ; la séparation peut être augmentée de moins de 30 pb sans aucun effet, mais de plus longues insertions peuvent inhiber l'initiation.

Un promoteur de l'ARNt, par conséquent, consiste en deux régions d'environ 20 pb chacune, qui doivent être séparées par plus de 20 pb et ne doivent pas être situées trop loin l'une de l'autre. Les séquences de ces régions sont hautement conservées dans les ARNt d'eucaryotes ; c'est un fait qui a été interprété uniquement en termes de fonction de l'ARNt, mais qui peut maintenant être attribué aux besoins du promoteur. (Le promoteur de l'ARNt est en rapport avec le promoteur de l'ARN 5S comme cela a été vu par les homologies entre les blocs A et l'extrémité 5' du promoteur de l'ARN 5S).

La localisation interne du promoteur pose une question importante. Lorsqu'un promoteur est situé à l'extérieur de l'unité transcriptionnelle elle-même, il peut probablement évoluer librement pour s'accorder aux besoins de l'enzyme. Mais les séquences nécessaires à l'initiation des unités de transcription par l'ARN polymérase III sont situées à l'intérieur de gènes de types plutôt différents, et sont par conséquent contraints de s'accorder aux besoins de produits aussi divers que l'ARN 5S, l'ARN VA (petits ARN transcrits à partir du génome de l'adénovirus), l'ARNt et les petits ARN nucléaires. Comment alors sont-ils capables de fournir aussi tous les caractères nécessaires à la reconnaissance par l'ARN polymérase III ?

Une solution à l'exigence pour l'enzyme de reconnaître une large variété de séquences de promoteurs est d'imaginer qu'elle pourrait agir par l'intermédiaire d'un facteur auxiliaire. Un facteur différent pourrait permettre à l'enzyme de se lier à chaque type de promoteur. Des facteurs spécifiques de gènes particuliers ont été trouvés dans différents systèmes.

**Figure 9.21**

Le facteur de contrôle auxiliaire du gène 5S se lie au brin non codant, il peut ainsi rester attaché à l'ADN pendant de nombreux cycles de transcription.

L'ARN polymérase III ne peut transcrire les gènes d'ARN 5S de *Xenopus* qu'en présence d'un facteur additionnel, une protéine de 37 000 daltons (*TFIIIA*) qui se lie à la région comprise entre + 45 et + 96. (La protéine sert à un double but ; elle se lie également au produit 5S dans l'oocyte.) Deux autres facteurs sont également nécessaires pour la transcription de tous les gènes de classe III, alors que la protéine de 37 000 daltons est spécifique des gènes 5S.

Ainsi, l'initiation peut dépendre de trois classes de protéines : l'ARN polymérase III elle-même, le(s) facteur(s) impliqué(s) généralement dans l'initiation, et le(s) facteur(s) spécifique(s) d'un ou de plusieurs gènes.

Les facteurs forment un **complexe de pré-initiation** sur le promoteur. Sur le promoteur du 5S de *Xenopus*, une simple copie de la protéine de 37 000 daltons est liée à la région de contrôle. Le complexe inclut aussi les deux autres facteurs.

Trois facteurs, A, B et C, sont également nécessaires pour initier la transcription de l'ARN 5S dans les cellules humaines. Mais avec les gènes ARNt, deux des facteurs, B et C, forment le complexe de pré-initiation ; avec le gène VA 1, le facteur C seul forme le complexe (bien que le facteur B reste nécessaire pour l'initiation). Chez la levure, deux facteurs, B et C, sont nécessaires pour transcrire les gènes d'ARNt ; mais des gènes individuels diffèrent dans leurs besoins de chaque facteur. Les deux blocs de séquences A et B sont nécessaires pour lier les facteurs, et leur séparation est également importante.

L'existence d'un complexe de pré-initiation signale que le gène est dans un état « actif », prêt à être transcrit. L'ARN polymérase peut se lier au gène seulement lorsque le(s) facteur(s) a (ont) été préalablement lié(s). Le complexe est stable et peut continuer d'exister pendant plusieurs cycles de réplication. La capacité à former un complexe de pré-initiation pourrait être un mécanisme général de régulation. En se liant à un promoteur, il rend possible la liaison de l'ARN polymérase ; le facteur, en fait, active le gène.

La **Figure 9.21** présente un modèle expliquant comment une molécule de facteur pourrait stimuler plusieurs cycles successifs de transcription à partir du promoteur. En restant fixé au brin non codant pendant le processus d'initiation, le facteur n'est pas déplacé. Nous pouvons supposer que d'autres facteurs de cette sorte sont encore à caractériser et prouveront leur importance dans la sélection de promoteurs.

Bien que la capacité de transcrire ces gènes soit conférée par le promoteur interne, le site d'initiation possède certaines influences. Des changements dans la région immédiatement en amont du site d'initiation peuvent altérer l'efficacité de transcription. Ainsi la cause primaire de reconnaissance réside dans le promoteur interne ; mais certaines responsabilités dans l'établissement de la fréquence d'initiation résident dans la région du site d'initiation.

POUR EN SAVOIR PLUS

Le volume édité par **Losick & Chamberlin**, *ARN polymerase* (Cold Spring Harbor Laboratory, New York, 1976) constitue une source de nombreuses revues et d'articles de recherche originaux. Deux chapitres couvrant le thème général traité ici ont été écrits par **Chamberlin**, ils en donnent tout d'abord une vision générale (pp. 17-68) puis présentent un exposé des interactions de l'enzyme bactérienne avec sa matrice (pp. 159-192). L'interaction moléculaire de l'ARN polymérase bactérienne avec ses promoteurs a été bien décrite par **Siebenlist, Simpson & Gilbert** (*Cell* 20, 269-281, 1980). Les étapes de l'interaction ont été traitées sous forme de revue par **Von Hippel et al.** (*Ann. Rev. Biochem.* 53, 389-449, 1984) et **McClure** (*Ann. Rev. Biochem.* 54, 171-204, 1985). L'initiation et son contrôle l'ont été par **Reznikoff et al.** (*Ann. Rev. Genet.* 19, 355-387, 1985). Les changements induits lors de la sporulation ont été traités sous forme de revue par **Losick & Pero** (*Cell* 25, 582-584, 1981).

Les techniques de caractérisation des promoteurs de l'ARN polymérase II d'eucaryotes ont été discutées par **Corden et al.** (*Science* 209, 1406-1414, 1981). La divergence entre la fonction du promoteur *in vivo* et *in vitro* a été éclairée par **McKnight et al.** (*Cell* 25, 385-398, 1981), qui a aussi étudié les composants du promoteur (*Science* 217, 316-324, 1982 ; *Cell* 31, 355-365, 1982), de même que **Dierks et al.** (*Cell* 32, 695-706, 1983). Les composants du promoteur et le facteur de liaison ont été traités sous forme de revue par **McKnight & Tjian** (*Cell* 46, 795-805, 1986). Une analyse approfondie sur le « enhancer » de SV40 a été réalisée par **Banerji, Rusconi & Schaffner** (*Cell* 27, 299-308, 1981) ; la régulation des « enhancers » par les récepteurs des stéroïdes a été passée en revue par **Yamamoto** (*Ann. Rev. Genet.* 19, 209-252, 1985). La transcription des gènes de 5S a été traitée sous forme de revue par **Korn** (*Nature* 295, 101-105, 1982). La nature des promoteurs dédoublés d'ARNt a été traitée par **Hall et al.** (*Cell* 29, 3-5, 1982).

CHAPITRE 10

UNE PANOPLIE D'OPÉRONS : LE MODÈLE DE L'OPÉRON LACTOSE ET D'AUTRES EXEMPLES

Les bactéries ont besoin de répondre rapidement aux changements de leur environnement. De brusques variations dans l'approvisionnement des aliments peuvent se produire à n'importe quel moment ; la survie dépend de la capacité à remplacer un substrat par un autre. Les eucaryotes unicellulaires peuvent aussi s'adapter à un environnement qui change sans cesse ; mais plus complexes, les organismes multicellulaires sont limités à un groupe de chaînes métaboliques plus constant et n'ont pas le même besoin de répondre aux variations extérieures.

La flexibilité est donc un avantage dans le monde bactérien. Cependant l'économie est également demandée, car une réponse aux exigences de l'environnement par des voies énergétiquement chères peut être désavantageuse. Il serait certainement coûteux de produire (d'une façon inutile) toutes les enzymes d'une voie métabolique qui ne sont pas utilisées car le substrat est absent. Ainsi le compromis bactérien est d'éviter la synthèse des enzymes d'une voie métabolique en l'absence de leur substrat et d'être prêt à produire ces enzymes si le substrat vient à apparaître.

Toutes ces contraintes expliquent les caractéristiques

principales de l'organisation des gènes bactériens. Ils ont tendance à s'organiser en groupes, les enzymes d'une voie déterminée étant représentées par des gènes adjacents. Le groupe de gènes en entier peut être transcrit en une seule espèce d'ARN messager polycistronique séquentiellement traduite par les ribosomes en chacune des protéines. Chez *E. coli* très peu de gènes sont individuellement transcrits, la majorité fait partie d'unités de transcription. Cette forme d'organisation permet à tous les gènes d'une unité transcriptionnelle d'être régulés d'une façon coordonnée grâce à l'interaction avec une protéine régulatrice se fixant à un site proche du promoteur.

Il semble quelquefois que tous les mécanismes concevables sont utilisés pour le contrôle de l'expression génétique dans une situation ou une autre. Mais à travers la variété des mécanismes de contrôle il s'en dégage un trait commun. La régulation résulte de l'interaction d'une macromolécule régulatrice avec une séquence d'acide nucléique, qui, souvent, possède une symétrie par rapport à un point (dyad symmetry). Très souvent la macromolécule est une protéine, mais dans certains cas, il s'agit d'un ARN. Il peut, également, s'agir d'un assemblage

macromoléculaire qui fonctionne par ailleurs dans la cellule pour une autre fonction. La base de ce type d'interaction est la capacité de la traduction par le ribosome à influencer la transcription par l'ARN polymérase.

L'INDUCTION ET LA RÉPRESSION SONT CONTRÔLÉES PAR DES PETITES MOLÉCULES

La synthèse des enzymes en réponse à l'apparition d'un substrat spécifique est appelée **induction**. Ce type de régulation est largement répandu chez les bactéries, il est également utilisé par des eucaryotes inférieurs (tels que les levures). Le système lactose d'*E. coli* fournit un modèle de ce type de mécanisme de contrôle.

Quand les cellules d'*E. coli* se développent en l'absence de β -galactosides, elles possèdent très peu de molécules de l'enzyme β -galactosidase, moins de cinq molécules par cellule. La fonction de cette enzyme est de couper les β -galactosides pour donner les sucres qui les composent. Par exemple le lactose est clivé en glucose et galactose qui seront ultérieurement métabolisés. Il n'y a pas besoin d'enzyme en l'absence de substrat.

Lorsque le substrat convenable est ajouté au milieu, l'activité enzymatique apparaît très rapidement dans la bactérie. Dans les 2 à 3 minutes un peu d'enzyme est présente, et très rapidement il peut y avoir plus de 5 000 molécules d'enzyme par bactérie. (Dans des conditions optimales la β -galactosidase peut représenter 5 à 10 % de la fraction protéique soluble d'une bactérie.) Si le substrat est éliminé du milieu, la synthèse de l'enzyme s'arrête aussi rapidement qu'elle avait démarré.

Ce type de réponse rapide aux changements de la composition du milieu en substances nutritives existe non seulement pour acquérir la capacité à métaboliser de nouveaux substrats mais également pour bloquer la synthèse endogène de composants qui peuvent soudainement apparaître dans le milieu. Par exemple *E. coli* synthétise l'acide aminé tryptophane grâce à l'enzyme tryptophane synthétase. Toutefois si le tryptophane est fourni dans le milieu de croissance, la production d'enzyme est immédiatement stoppée. Cet effet est appelé **répression**. Il empêche la bactérie de gaspiller ses ressources à des activités synthétiques non nécessaires.

L'induction et la répression représentent le même phénomène. Dans un cas la bactérie adapte sa capacité à utiliser un substrat donné pour sa croissance, dans l'autre cas elle adapte sa capacité à synthétiser un

intermédiaire métabolique particulier. Le déclencheur de chaque type d'ajustement est, soit la petite molécule substrat de l'enzyme, soit le produit de l'activité enzymatique. Les petites molécules capables de produire les enzymes qui les métabolisent sont appelées **inducteurs**. Celles qui empêchent la production des enzymes capables de les synthétiser sont appelées **corépresseurs**.

La capacité d'agir comme un inducteur ou comme un corépresseur est très spécifique. Seuls le substrat, le produit ou les molécules très proches sont actifs. Mais dans les deux cas, *l'activité de la petite molécule ne dépend pas de son interaction avec l'enzyme cible*. Certains inducteurs ressemblent aux inducteurs naturels de la β -galactosidase mais ne peuvent pas être métabolisés par l'enzyme. L'exemple par excellence est l'isopropylthiogalactoside (IPTG), un des nombreux thiogalactosides possédant cette propriété. Quoiqu'il ne soit pas reconnu par la β -galactosidase, l'IPTG est néanmoins un inducteur très puissant.

Les molécules qui induisent la synthèse enzymatique mais qui ne sont pas métabolisées sont appelées **inducteurs gratuits**. Ils sont très utiles car ils se maintiennent dans la cellule sous leur forme d'origine. (Un véritable inducteur devrait être métabolisé, interférant ainsi avec l'étude du système.) L'existence d'inducteurs gratuits suggère un point important. *Le système doit posséder certains composants distincts de l'enzyme cible, reconnaissant le substrat approprié : la capacité de ces composants à reconnaître des substrats potentiels de la même famille est différente de celle de l'enzyme.*

LES GROUPES DE GÈNES STRUCTURAUX SONT RÉGULÉS D'UNE FAÇON COORDONNÉE

En général, toutes les enzymes d'une chaîne métabolique sont régulées simultanément par induction ou par répression. En plus des enzymes directement impliquées dans la chaîne métabolique, d'autres activités peuvent être incluses dans l'unité de régulation coordonnée : par exemple la protéine responsable du transport de la petite molécule substrat à l'intérieur de la cellule.

Les gènes qui codent pour des protéines requises par la cellule pour des activités enzymatiques ou des fonctions structurales sont appelés des **gènes de structure**. La grande majorité des gènes bactériens rentre dans cette catégorie, qui représente ainsi une énorme variété de structures protéiques et de fonctions.

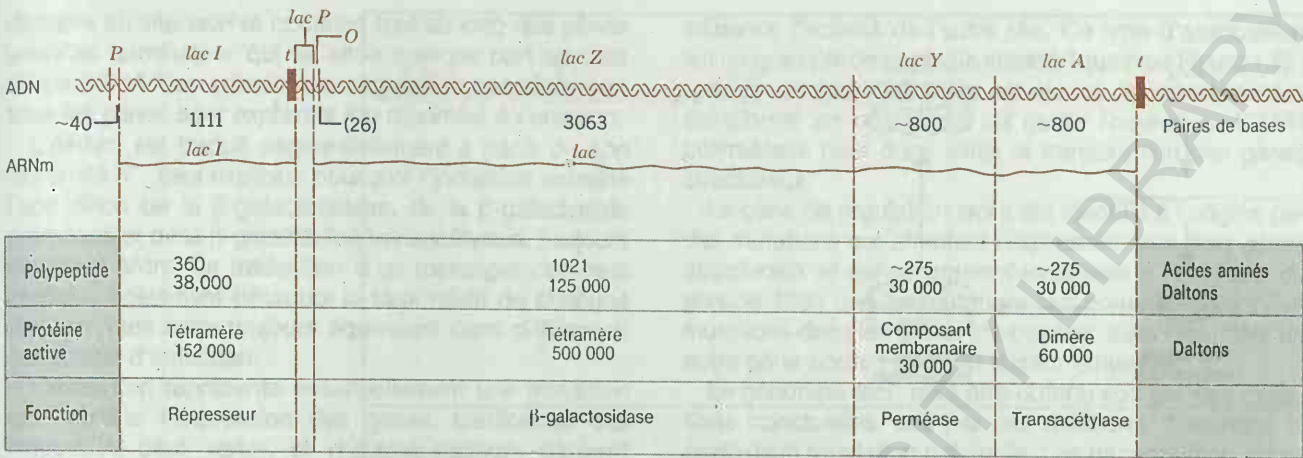


Figure 10.1
L'opéron lactose occupe environ 6 000 pb d'ADN.
A gauche, le gène *lac I* possède son propre promoteur et terminateur. La fin de la région *lac I* est adjacente au promoteur *lac P*. L'opérateur *lac O* occupe les 26 premières paires de bases du gène *lac Z*, qui est extrêmement long et suivi des gènes *lac Y*, *lac A*, ainsi que d'un terminateur.

Les gènes structuraux comprennent également les gènes codant pour les ARN ribosomaux (ARNr) et de transfert (ARNt). Tous les gènes structuraux tendent à s'organiser en groupes qui peuvent être contrôlés d'une manière coordonnée, comme par exemple les trois gènes de structure *lac*, *lac Z*, *Y* et *A*.

La carte du groupe est dessinée **Figure 10.1** :

le gène *lacZ* code pour l'enzyme β-galactosidase, dont la forme active est un tétramère d'environ 500 000 daltons ;

le gène *lacY* code pour la β-galactoside perméase, une protéine membranaire de 30 000 daltons qui constitue le système de transport ;

le gène *lacA* code pour la β-galactoside transacétylase, une enzyme qui transfère le groupe acétyl de l'acétyl-CoA aux β-galactosides.

Des mutations, soit dans le gène *lacZ*, soit dans le gène *lacY*, entraînent le génotype *lac⁻*, les cellules ne peuvent pas utiliser le lactose. Les mutations *lacZ⁻* abolissent l'activité enzymatique, empêchant le métabolisme du lactose. Le lactose du milieu ne peut pas être introduit dans les cellules chez les mutants *lacY⁻*. Aucun défaut n'a été identifié dans les cellules *lacA⁻*, ce qui reste énigmatique. (Lorsque la bactérie croît en présence d'un analogue non métabolisable de β-galactosides, il est possible que la réaction d'acétylation puisse donner un

avantage sélectif, car celle-ci entraîne la détoxification et l'excrétion du produit).

La cible de la petite molécule inductrice est une protéine spéciale qui possède comme seule fonction le contrôle de l'expression des gènes structuraux. Le gène qui code pour cette protéine est appelé **gène régulateur**. Les gènes régulateurs sont responsables du contrôle de l'expression des groupes de gènes structuraux. Un gène régulateur code, en général, pour une protéine régulatrice qui contrôle la transcription en se fixant sur des sites spécifiques de l'ADN.

Nous pouvons distinguer les gènes structuraux des gènes régulateurs par l'effet des mutations. Une mutation dans un gène structural prive la cellule de la protéine codée par ce gène. Au contraire une mutation dans un gène régulateur modifie l'expression de tous les gènes structuraux qu'il contrôle. La nature de cette modification révèle le type de régulation exercé.

Les gènes *lac* sont contrôlés par une **régulation négative** : ils sont transcrits jusqu'à ce qu'ils soient inactivés par la protéine régulatrice. Une mutation qui inactive le régulateur entraîne l'expression constitutive des gènes. Puisque la fonction du régulateur est d'empêcher l'expression des gènes structuraux, il est appelé **répresseur**.

Le concept que deux classes de gènes pouvaient être distinguées par leurs fonctions a été proposé par Jacob et Monod en 1961 dans la formulation classique du

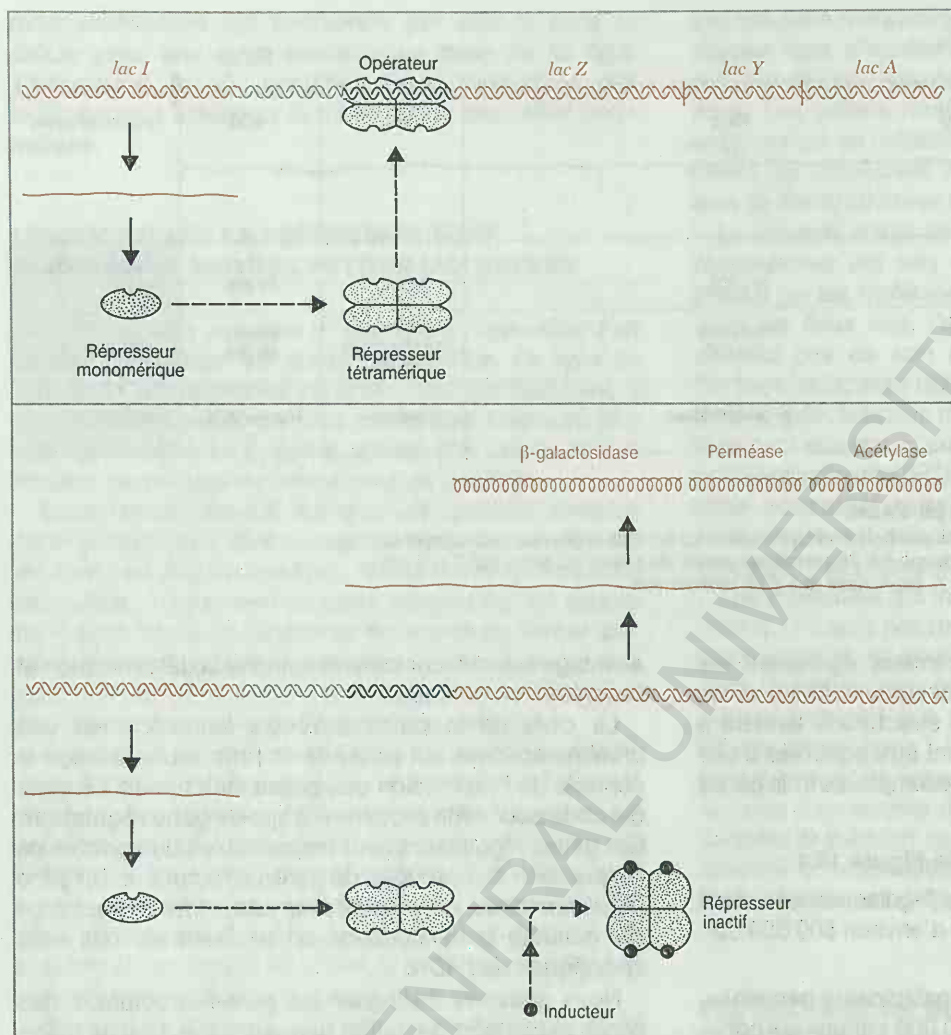


Figure 10.2

L'opéron lactose est activé par un inducteur.

En haut : Le gène *lac I* synthétise un répresseur sous forme de tétramère qui se fixe à l'opérateur en empêchant la transcription des gènes structuraux.

En bas : L'addition de l'inducteur transforme le répresseur en une forme inactive qui ne peut plus se lier à l'opérateur ; ainsi la transcription démarre au promoteur et les trois enzymes sont synthétisées.

modèle de l'opéron. L'**opéron** est une unité d'expression génétique qui inclut des gènes structuraux et les éléments contrôlant leur expression. L'activité de l'opéron est contrôlée par des gènes régulateurs codant pour des protéines interagissant avec les éléments de contrôle.

La régulation de base de l'opéron lactose est illustrée dans la **Figure 10.2**. Le groupe de trois gènes *lac Z*, *Y*

et *A* est transcrit en un seul message à partir du promoteur situé en amont du gène *lac Z*. La capacité à initier la transcription au promoteur *lac P* est contrôlée par la protéine répresseur codée par le gène régulateur *lac I*.

En l'absence d'inducteur le groupe de gènes n'est pas transcrit. Lorsqu'un inducteur est ajouté, la transcription

démarré au site *lacP* et continue tout au long des gènes jusqu'au terminateur qui se situe quelque part au-delà du gène *lacA*. Ceci entraîne une **régulation coordonnée** : tous les gènes sont exprimés (ou réprimés) à l'unisson.

L'ARNm est traduit séquentiellement à partir de son extrémité 5' ; ceci explique pourquoi l'induction entraîne l'apparition de la β -galactosidase, de la β -galactoside perméase et de la β -galactoside transacétylase, toujours dans cet ordre. La traduction d'un messager commun explique également pourquoi le taux relatif de chacune des enzymes reste toujours équivalent dans différentes conditions d'induction.

L'induction représente essentiellement une transition qui entraîne l'expression des gènes. L'efficacité des inducteurs peut varier, et d'autres facteurs peuvent influencer le niveau absolu de transcription ou de traduction, mais le rapport entre les trois gènes est prédéterminé par leur organisation.

Le messager *lac* est extrêmement instable, son temps de demi-vie est de 3 minutes seulement. Cette caractéristique permet à l'induction d'être arrêtée très rapidement car en très peu de temps, dès que l'inducteur disparaît, la transcription est stoppée et tous les ARNm lactose sont détruits ; la cellule bloque ainsi la production des enzymes.

LA PROTÉINE RÉPRESSEUR SE FIXE AU NIVEAU DE L'OPÉRATEUR

Le répresseur empêche la transcription par sa fixation sur une séquence d'ADN appelée **opérateur** : *lacO*. Ce site s'étend entre le promoteur (*lacP*) et le groupe des gènes structuraux (*lac Z*, *Y* et *A*). Lorsque le répresseur se fixe à l'opérateur, sa présence empêche l'ARN polymérase d'initier la transcription au promoteur.

Comment cette interaction répond-elle à la petite molécule inductrice ? La protéine répresseur possède une très forte affinité pour l'opérateur ; en l'absence d'inducteur, il se fixe de telle façon que les gènes structuraux adjacents ne peuvent pas être transcrits. Toutefois, l'inducteur peut s'associer au répresseur pour former un complexe répresseur-inducteur qui ne peut plus se lier à l'opérateur.

Les caractéristiques essentielles du contrôle du système résident dans la double propriété du répresseur : il peut empêcher la transcription et reconnaître la petite molécule inductrice. Le répresseur a deux sites d'interaction, un pour l'inducteur et un autre pour l'opérateur. Quand l'inducteur se fixe sur son site, il change la conformation de la protéine de telle manière qu'il

influence l'activité de l'autre site. Ce type d'association est un exemple de **contrôle allostérique** (voir Figure 1.6).

Ainsi lorsque l'inducteur est ajouté, le répresseur est transformé en une forme qui quitte l'opérateur. L'ARN polymérase peut donc initier la transcription des gènes structuraux.

Le gène de régulation *lacI* a été identifié à l'origine par des mutations qui affectent l'expression des trois gènes structuraux et qui se trouvent localisées à l'extérieur du groupe. Bien que les mutations *lacI* complémentent des mutations dans les gènes structuraux, elles identifient un autre gène codant pour un facteur diffusible.

Le génotype *lacI⁻* peut être obtenu soit par des mutations ponctuelles, soit par des délétions. Il entraîne la perte de la régulation habituelle. Les mutants *lacI⁻* expriment les gènes structuraux tout le temps, indépendamment de la présence ou de l'absence de l'inducteur. Cette induction est décrite comme une expression génétique **constitutive**.

Un tel comportement est conforme à notre attente dans un système de régulation négatif. Le gène *lacI⁺* code pour une protéine répresseur qui est capable d'empêcher la transcription du groupe de gènes *lac Z*, *Y* et *A*. Une mutation entraînant le type *lacI⁻* permet aux gènes d'être constitutivement exprimés car le répresseur devient inactif.

Nous pouvons renforcer cette conclusion en déterminant le rapport entre le gène mutant constitutif *lacI⁻* et le type sauvage *lacI⁺* lorsque les deux sont présents dans la même cellule. Ceci est obtenu avec la formation d'un **diploïde partiel** : une copie de l'opéron est présente dans le génome bactérien, et l'autre est introduite par l'intermédiaire d'un **plasmide**, une molécule d'ADN indépendante autorépliquative qui possède seulement un petit nombre de gènes, dont une copie totale ou partielle de l'opéron.

Les cellules dans lesquelles un gène régulateur est *lacI⁺* et l'autre *lacI⁻* possèdent une régulation normale. Les gènes structuraux sont à nouveau réprimés quand l'inducteur est retiré du milieu. Nous pouvons expliquer cet effet, comme le montre la Figure 10.3.

Les mutations constitutives dans le répresseur détruisent sa capacité d'interagir avec l'opérateur. Ainsi la transcription peut s'initier librement au promoteur. Mais l'introduction d'un second gène régulateur de type sauvage rétablit la présence d'un répresseur normal. Ainsi, à nouveau, l'opéron est inactif en l'absence d'inducteur. En termes de génétique l'**inductibilité** de type sauvage est **dominante** sur la mutation constitutive. Ceci est la caractéristique principale d'un contrôle négatif.

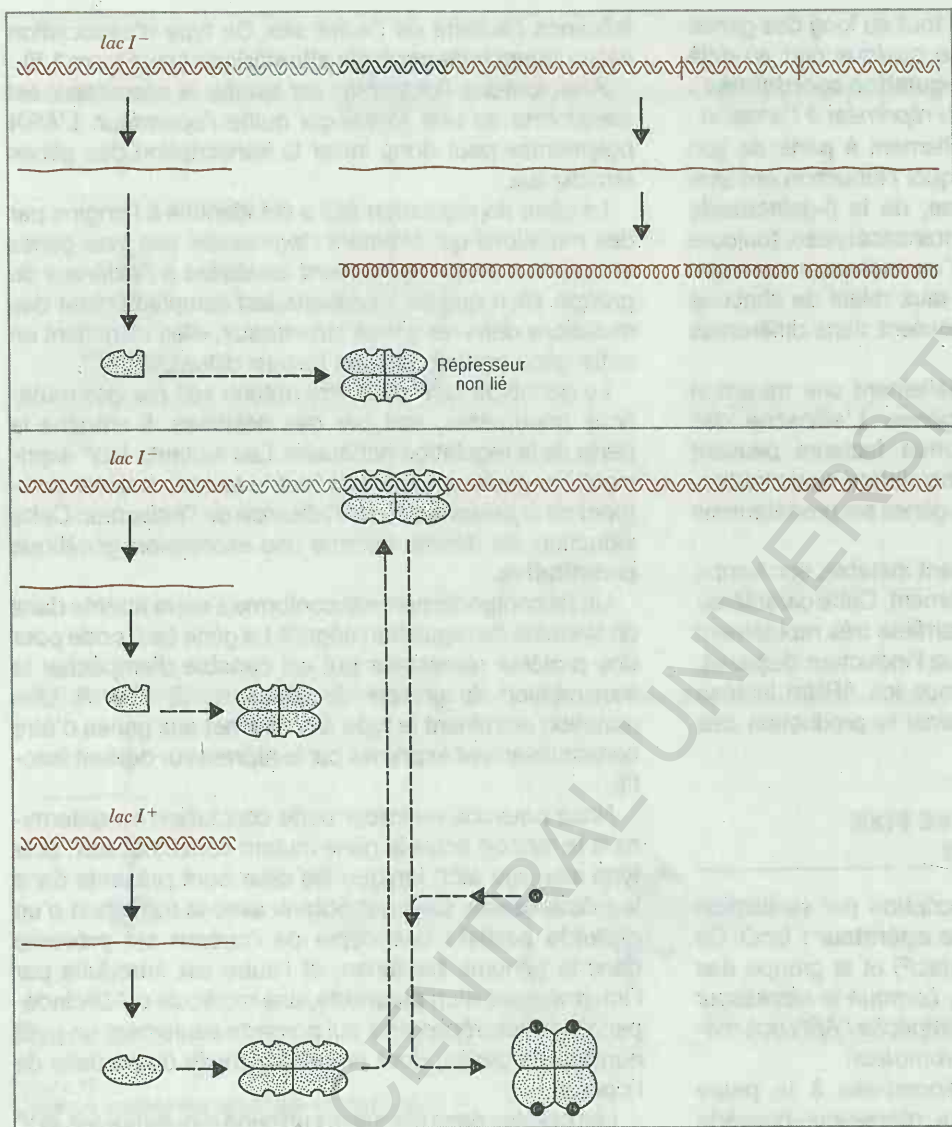


Figure 10.3
Les mutants constitutifs dans le gène *lacI* sont récessifs.

En haut : La mutation dans le gène *lacI* rend le répresseur inactif, car il ne peut plus se lier à l'opérateur ; les gènes structuraux s'expriment.

En bas : L'introduction d'un deuxième gène *lacI⁺* permet la production d'un répresseur normal. Il est capable d'agir en *trans* et de se fixer à l'opérateur. Il peut être déplacé de l'ADN par l'inducteur lorsque celui-ci est ajouté. Ce qui rétablit la régulation normale des gènes structuraux.

Le répresseur lactose est un tétramère de sous-unités identiques de 38 000 daltons chacune. Il y a moins de 10 tétramères par cellule (ce qui signifie qu'il doit y avoir 40 traductions du gène *lacI* par cycle cellulaire). Le gène régulateur est transcrit en un ARN messager monocistronique à une vitesse qui paraît être uniquement gouvernée par l'affinité de l'ARN polymérase pour son promoteur.

Le gène *lacI* se trouve à gauche du groupe de gènes structuraux et forme une unité indépendante de transcrip-

tion. Comme le gène *lacI* code pour un produit diffusible, actif en *trans*, il n'a en principe pas besoin d'être localisé près des gènes structuraux. En effet comme nous l'avons vu un gène *lacI* situé sur un plasmide indépendant est capable de contrôler l'expression du groupe *lac Z, Y et A* du chromosome bactérien.

Dans d'autres opérons d'*E. coli*, le gène régulateur est localisé à une certaine distance des gènes structuraux cibles. Nous pouvons toutefois faire l'hypothèse qu'un

avantage de conserver un gène régulateur proche de ses gènes de structure est qu'il n'a une utilité qu'en association avec eux.

L'OPÉRATEUR AGIT EN *CIS*

L'opérateur a été initialement identifié grâce à des mutations constitutives dénommées o^c dont les propriétés particulières démontrent que *cette région ne représente pas un produit diffusible*.

Les gènes structuraux contigus à une mutation o^c sont exprimés constitutivement. La mutation change de telle sorte l'opérateur que le répresseur ne peut plus interagir avec lui. Ainsi le répresseur n'empêche plus l'ARN polymérase d'initier la transcription. L'opéron est donc continuellement exprimé, comme c'est illustré dans la Figure 10.4.

L'opérateur peut contrôler uniquement les gènes *lac* qui lui sont adjacents. Si un deuxième opéron lactose est introduit dans la bactérie sous forme d'une molécule d'ADN indépendante, possédant son propre promoteur, aucun opérateur n'est influencé par l'autre. Ainsi si un opéron possède un opérateur de type sauvage, il sera réprimé dans les conditions habituelles ; au contraire le second opéron avec une mutation o^c est exprimé de manière constitutive.

La capacité d'un site à contrôler les gènes adjacents indépendamment de la présence dans la cellule d'autres allèles du site en question définit le phénomène de **dominance en *cis***. Ceci indique que le site muté ne code pour aucun produit (protéine, en général, mais aussi, en principe, ARN) qui puisse diffuser dans la cellule pour exercer son effet sur un autre allèle. Souvent un site *cis*-dominant est dit **actif en *cis*** pour contraster avec la fonction de ***trans*-activation** qui, parce qu'elle corres-

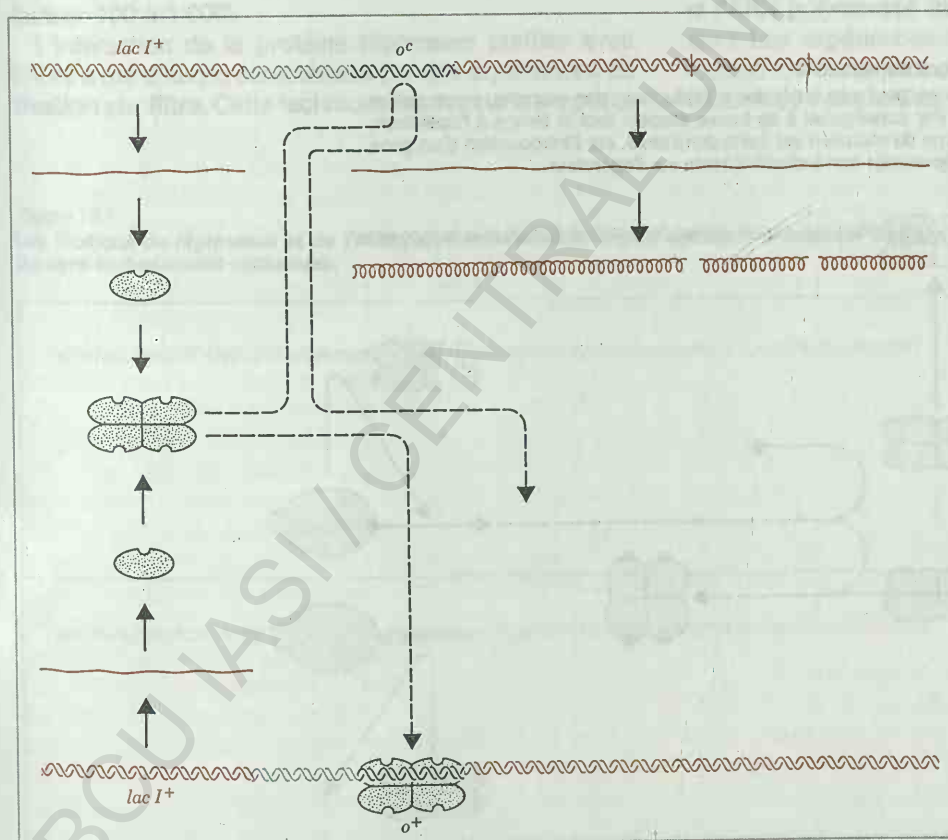


Figure 10.4

Les mutations constitutives de l'opérateur (o^c) sont *cis*-dominantes.

En haut : L'opéron possède un opérateur muté qui ne peut pas fixer le répresseur, ainsi les gènes structuraux adjacents sont exprimés constitutivement.

En bas : L'opéron possède un opérateur de type sauvage, qui répond au répresseur et s'exprime seulement au cours de l'induction.

Le comportement de chaque groupe de gènes structuraux est contrôlé uniquement par l'opérateur contigu ; ses propriétés ne sont jamais influencées par un autre opérateur.

pond à un produit diffusible, peut agir au niveau de tous les sites appropriés dans la cellule, qu'ils soient présents ou pas dans la même molécule d'ADN (voir Chapitre 1).

Le concept de la dominance en *cis* est appliqué à *n'importe quelle séquence d'ADN dont la fonction est d'être « reconnue par » plutôt que de « coder pour » un produit diffusible*. Ainsi les mutations dans un promoteur ou dans un terminateur agissent-elles également en *cis*.

Une mutation d'un site agissant en *cis* ne peut pas être assignée à un groupe de complémentation. (La capacité de complémentation est exclusivement caractéristique des gènes s'exprimant sous la forme de produits diffusibles). Lorsque deux sites agissant en *cis* sont proches, par exemple un promoteur et un opérateur, nous ne pouvons pas classer les mutations par un test de complémentation. Nous sommes donc limités à les distinguer par leurs effets sur le phénotype.

La dominance en *cis* est une caractéristique d'un site quelconque qui est *physiquement lié à la séquence qu'il contrôle*. Si le site de contrôle fonctionne comme une partie de l'ARN messager polycistronique, des mutations

à l'intérieur montreraient *exactement le même profil* de « *cis*-dominance » que si elles fonctionnaient dans l'ADN. Le point critique est que le site de contrôle ne peut être physiquement séparé des gènes qu'il contrôle. Du point de vue génétique, peu importe si le site et les gènes se trouvent ensemble sur l'ADN ou sur l'ARN.

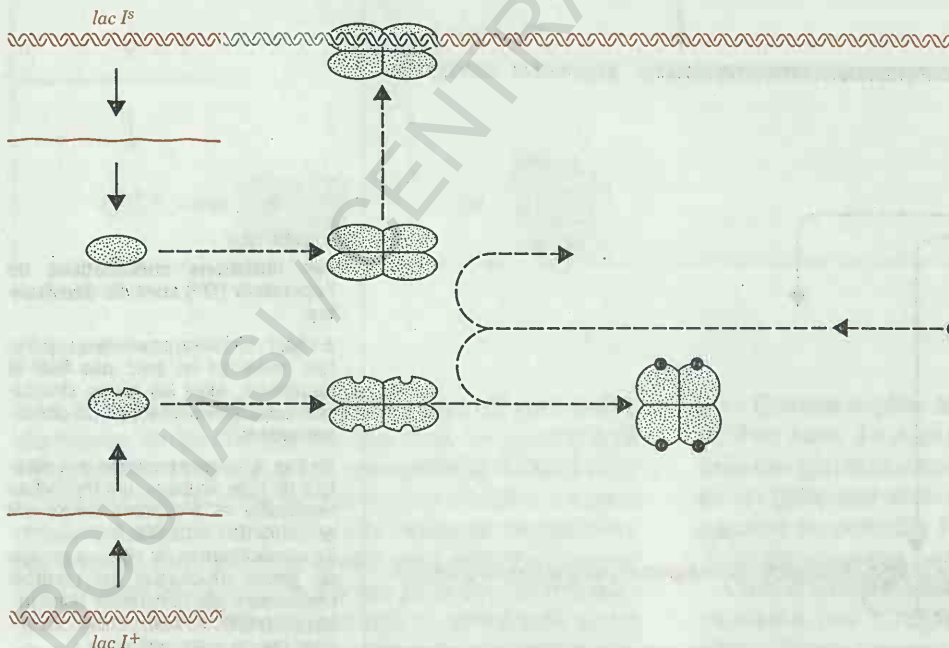
Les mutants de l'opéron qui sont **non inductibles** ne peuvent pas être exprimés du tout. Ils peuvent, comme les mutants constitutifs, être rangés en deux catégories génétiques. Les rapports de dominance de chaque type peuvent être là aussi utilisés pour définir la nature du locus :

- Les mutations du promoteur agissent en *cis* ; elles empêchent l'ARN polymérase d'interagir avec le promoteur *lacP*, et rendent ainsi l'opéron non fonctionnel car il ne peut pas être transcrit.
- Les mutations dans le gène *lacI* qui abolissent la capacité du répresseur à s'associer à l'inducteur peuvent entraîner le même phénotype. Le répresseur est bloqué sous sa forme active qui lie l'opérateur et empêche la transcription.

Figure 10.5

Les mutations *lacI^s* non inductibles sont dominantes.

Le gène *lacI^s* produit un répresseur qui ne peut pas répondre à l'inducteur, par exemple, parce qu'il a perdu le site de fixation de l'inducteur. Par conséquent il se trouve associé tout le temps à l'opérateur, même en présence de l'inducteur. Ce type de mutation est *trans* dominante, car l'introduction d'un gène *lacI* n'entraîne aucun changement ; le répresseur non inductible reste sur l'opérateur.



L'addition de l'inducteur est sans résultat. De tels mutants sont désignés sous le nom de *lac^s*. La **Figure 10.5** montre qu'ils sont *trans*-dominants sur le type sauvage. Ceci se produit car le répresseur muté se lie à tous les opérateurs de la cellule en empêchant leur transcription et il ne peut être délogé, quelles que soient les propriétés de n'importe quelle autre protéine répresseur présente.

COMMENT LE RÉPRESSEUR BLOQUE-T-IL LA TRANSCRIPTION ?

Le répresseur a été initialement isolé en purifiant le composant d'extraits d'*E. coli* capable d'interagir avec l'inducteur gratuit IPTG. (Les quantités de répresseur dans la cellule étant très faibles, il a été indispensable pour obtenir assez de matériel, d'utiliser une mutation « up » du promoteur afin d'augmenter la transcription de *lacI*. Ce locus *lacI* muté a été ensuite placé en de multiples copies par cellule. La synthèse est ainsi augmentée d'un facteur 100 à 1 000).

L'interaction de la protéine répresseur purifiée avec l'ADN a été d'abord caractérisée par des expériences de **fixation sur filtre**. Cette technique repose sur la propriété

des filtres de nitrocellulose de retenir des protéines mais non de l'ADN double brin. L'ADN complexé à des protéines est également retenu. Ainsi les séquences qui peuvent se lier spécifiquement au répresseur sont retenues par le filtre.

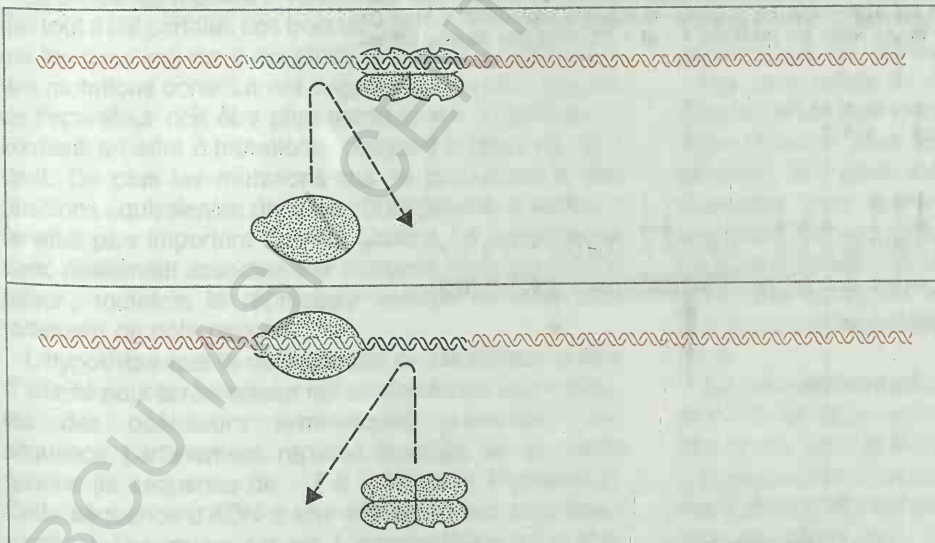
Le répresseur se fixe à l'ADN contenant la séquence de l'opérateur *lac* de type sauvage. L'ADN doit être sous la forme double-brin. Le répresseur n'interagit pas avec l'ADN obtenu d'un mutant *lac^s*. L'addition d'IPTG libère le répresseur de l'ADN *lacO* *in vitro*. La réaction entre le répresseur et l'ADN *lacO* *in vitro* possède les mêmes caractéristiques de contrôle qu'*in vivo*, si bien qu'il peut être utilisé pour établir les bases de la répression.

Quelle est la localisation précise de l'opérateur par rapport au promoteur ? L'opérateur *lacO* comprend une région d'environ 26 pb qui est définie par la longueur d'ADN protégée par le répresseur de l'attaque *in vitro* par des nucléases. Il s'étend de la position - 5 en amont du début de l'ARN messager jusqu'à la position + 21 qui se trouve dans l'unité de transcription. Ainsi recouvre-t-il l'extrémité droite du promoteur.

Est-ce le chevauchement qui empêche le répresseur et l'ARN polymérase de se fixer simultanément à l'ADN *lac* ? Les expériences de compétition montrent que la fixation de l'une des protéines sur l'ADN empêche la fixation de l'autre. Ceci est illustré dans la **Figure 10.6**.

Figure 10.6

Les fixations du répresseur et de l'ARN polymérase dans la région opérateur-promoteur du gène *lac* sont mutuellement exclusives.



Des résultats similaires ont été obtenus dans d'autres opérons, bien que l'opérateur ne chevauche pas toujours la même région du promoteur (voir la Figure 10.13). Par exemple dans le phage lambda, l'opérateur s'étend en amont de la région promotrice (voir Chapitre 12). Cependant la règle générale est que *lorsque le répresseur est lié à son opérateur, sa présence bloque l'accès de l'ARN polymérase au promoteur en empêchant ainsi l'initiation de la transcription.*

LES CONTACTS DANS L'OPÉRATEUR

Comment le répresseur reconnaît-il la séquence d'ADN spécifique de l'opérateur ? L'interaction entre le répresseur et l'opérateur est souvent choisie comme modèle d'interaction spécifique de protéines avec l'ADN.

L'opérateur possède un trait commun avec plusieurs

sites de reconnaissance de protéines régulatrices : il possède un axe de symétrie « dyad » (c'est-à-dire par rapport à un point). L'inversion répétée est présentée dans la Figure 10.7.

Les approches utilisées pour l'analyse de l'interaction polymérase-promoteur (voir Chapitre 9), sont les mêmes que celles employées pour définir les points de contact du répresseur et de l'opérateur. Les mutations ponctuelles constitutives permettent d'identifier des paires de bases, essentielles pour l'interaction ; des délétions permettent de délimiter les bords de la région impliquée. Lorsque l'ADN lié au répresseur est comparé avec l'ADN non lié par les expériences de méthylation ou de fixation covalente aux UV, on peut identifier les bases protégées, ou plus sensibles, lorsqu'elles sont associées à la protéine.

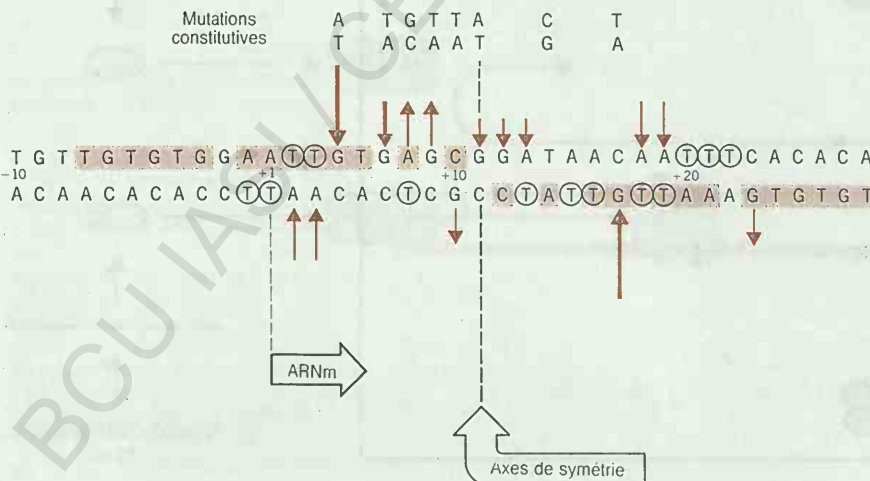
La région de symétrie dans l'opérateur s'étend de la position -8 à la position +28 ; cependant la région de l'ADN protégée de l'attaque des nucléases par la fixation

Figure 10.7
L'opérateur lactose possède une séquence symétrique.

La séquence est numérotée +1 au point d'initiation de la transcription. Les régions de symétrie opposée sont indiquées par les zones ombrées, l'axe de symétrie passe par la paire de bases à la position +11.

Les substitutions qui entraînent une expression constitutive sont indiquées par les paires de bases au-dessus de la séquence. Les deux mutations à droite sont exactement les symétriques des mutations trouvées à gauche. A la position +17, T-A remplace G-C, alors qu'à la position +5, A-T remplace G-C. De la même manière à la position +14, C-G remplace T-A, alors qu'à la position +8, G-C remplace A-T.

Les thymines aux positions dans lesquelles la substitution par le bromo-uracile permet la liaison covalente du répresseur sont indiquées par un cercle. Les flèches pointées vers la séquence indiquent les purines qui sont protégées par le répresseur d'une méthylation. Les flèches dirigées vers l'extérieur de la séquence identifient les positions où la méthylation est augmentée. La longueur de la flèche représente la force de l'effet. La région de plus fort contact se trouve entre les positions +1 et +23, mais des liaisons faibles peuvent s'établir d'un côté ou de l'autre.



du répresseur est plus courte et comprend 26 paires de bases entre les positions - 5 et + 21. La région identifiée par les mutations constitutives est même plus petite. Elle se trouve à l'intérieur de la région centrale, et s'étend sur 13 pb de la position + 5 à la position + 17 ; il existe 8 sites au niveau desquels la substitution d'une seule base entraîne une constitutivité. Ceci confirme le fait discuté pour les mutations des promoteurs et qui a été résumé dans la Figure 9.11. *Un minimum de contacts essentiels dans une large région peuvent être responsables de l'association d'une protéine avec la séquence d'ADN spécifique.*

Les exemples d'activation et de protection des bases montrent certains traits de symétrie caractéristiques à l'intérieur de la région de contacts forts. La Figure 10.7 montre que seuls deux résidus thymines dans la région + 1 à + 22 peuvent se lier au répresseur. Les expériences de méthylation montrent qu'une large proportion des purines (A et G) entre les positions + 3 à + 19 est protégée assez fortement par le répresseur. Un petit nombre présente une sensibilité accrue, probablement dû à la création de poches hydrophobes dans le complexe répresseur-opérateur.

Les interactions avec les positions + 1 et + 6 sont presque symétriques de celles qui se produisent aux positions + 21 et + 16 ; les contacts les plus proches du point de symétrie ne sont pas symétriques. Ainsi le répresseur se lie à l'opérateur de manière à se placer symétriquement au niveau des points de contact périphériques, alors que la liaison ne présente pas de symétrie au voisinage immédiat du centre de symétrie.

La séquence répétée inversée de l'opérateur *lac* n'est pas tout à fait parfaite. Les trois différences sont montrées par les cassures dans les blocs ombrés. La distribution des mutations constitutives suggère que le côté gauche de l'opérateur doit être plus sensible aux mutations : il contient en effet 6 mutations, comparé à deux du côté droit. De plus les mutations qui se produisent à des positions équivalentes dans les côtés gauche et droit ont un effet plus important du côté gauche. La symétrie est donc réellement associée aux contacts répresseur-opérateur ; toutefois le répresseur semble se fixer plus fortement du côté gauche.

L'hypothèse que le côté gauche de l'opérateur a plus d'affinité pour le répresseur est confirmée par les propriétés des opérateurs synthétiques présentant une séquence parfaitement répétée inversée de la moitié gauche (la séquence de - 7 à + 10 de la Figure 10.7). Cette séquence d'ADN a une affinité 10 fois supérieure à celle de l'opérateur naturel. L'augmentation est probablement due à la transformation de la moitié droite en

l'image exacte de la moitié gauche, ainsi qu'à l'élimination de la paire de bases centrale (+ 11) qui rapproche les deux moitiés.

La symétrie de l'ADN reflète-t-elle la symétrie de la protéine ? C'est probablement le cas, car le répresseur est un tétramère de sous-unités identiques, qui doivent certainement posséder chacune le même site de fixation pour l'ADN. Chaque moitié de l'opérateur est probablement associée de la même manière à un dimère du répresseur.

La région centrale essentielle pour la fixation du répresseur occupe grossièrement les 20 premières bases en aval du site d'initiation de la transcription. Cependant, si nous nous rappelons les points de contacts de l'ARN polymérase avec le promoteur (Figure 9.11), ils se situent en amont du site d'initiation. Ainsi les points de contacts majeurs pour les deux protéines sont adjacents plutôt que chevauchants.

LE RÉPRESSEUR EST UNE PROTÉINE MULTIMÉRIQUE SE LIANT À L'ADN

L'interaction entre le répresseur et ses deux types de sites de liaison contrôle l'expression génétique en réponse à l'environnement. Le *site de fixation* à l'ADN reconnaît la séquence de l'opérateur. Le *site de fixation de l'inducteur* fixe la petite molécule inductrice, et il résulte de cette interaction que les sites de fixation pour l'ADN perdent leur capacité de liaison pour l'opérateur.

Les deux sites de fixation peuvent être identifiés dans la sous-unité du répresseur par des mutations *lacI* qui les inactivent. Leur relation *in vivo* dépend de la structure multimérique du répresseur.

Les sous-unités du répresseur s'associent au hasard dans la cellule pour constituer la protéine active tétramérique. Lorsque deux allèles différents du gène *lacI* sont présents, les sous-unités issues de chacun peuvent s'associer pour former un hétérotétramère, dont les propriétés peuvent différer de celles des homotétramères. Ce type d'interaction entre les sous-unités est une propriété des protéines multimériques et elle est décrite comme une **complémentation interallélique** (voir Chapitre 1).

La **complémentation négative** existe entre certains mutants de répresseurs, comme dans la combinaison des gènes *lacI^d* et *lacI⁺*. La mutation *lacI^d* seule conduit à la production d'un répresseur qui ne peut pas se lier à l'opérateur et elle est par conséquent constitutive comme tous les allèles *lacI⁺*. Les mutations du type *lacI⁻* sont récessives par rapport au type sauvage, car elles inacti-

vent le répresseur. Toutefois la notation $-d$ indique que ce variant de type négatif est dominant lorsqu'il est couplé à un allèle de type sauvage.

La cause de la dominance est que l'allèle $lacI^d$ produit une sous-unité défectueuse qui est non seulement incapable de se lier à l'opérateur, mais qui de plus empêche n'importe quelle sous-unité normale du tétramère d'interagir avec l'opérateur. Ceci démontre que la répression est obtenue par l'intermédiaire du tétramère plutôt que par celle du monomère. L'effet d'empoisonnement peut également être produit *in vitro* par le mélange approprié de « bonnes » et de « mauvaises » sous-unités.

Les mutations $lacI^d$ permettent d'identifier le site de fixation à l'ADN dans la sous-unité du répresseur. Ceci explique leur capacité à empêcher les tétramères mixtes de se lier à l'opérateur. Une réduction dans le nombre des sites de fixation doit énormément réduire l'affinité spécifique. La carte du gène $lacI$ est présentée dans la **Figure 10.8** : les mutations $lacI^d$ sont groupées à l'extrémité gauche du gène. Ceci permet d'identifier la région N-terminale de la protéine comme un site de fixation à l'ADN. Les mutations récessives de type $lacI^-$ peuvent se produire partout ailleurs dans la molécule et peuvent exercer indirectement leurs effets dans l'interaction avec l'ADN.

Le rôle de la région N-terminale dans la fixation spécifique avec l'ADN est également confirmé par la localisation à son niveau des mutations « liaison étroite ». Ces mutations qui sont rares augmentent l'affinité du répresseur pour l'opérateur ; l'interaction est souvent si forte qu'elle n'est pas rompue par l'inducteur.

Les mutations non inductibles de type $lacI^s$ rendent le

répresseur insensible à l'inducteur. Ceci est obtenu soit parce que la protéine a perdu son site de fixation pour l'inducteur, soit parce qu'elle devient incapable de transmettre au site de fixation de l'ADN l'information issue du site de fixation de l'inducteur. Comme il est montré dans la Figure 10.8, les mutations $lacI^s$ sont groupées et espacées régulièrement le long du gène. Les espaces peuvent représenter des tours dans la chaîne polypeptidique.

Le comportement de la protéine répresseur isolée *in vitro* montre clairement qu'elle possède un site de fixation pour l'ADN dont la capacité à rester liée à l'opérateur est influencée par la structure du reste de la molécule.

Quand le répresseur est traité par la trypsine, il est coupé préférentiellement à l'acide aminé 59. La région C-terminale de la protéine avec les résidus 60-360 est décrite comme étant le **cœur résistant à la trypsine**. Il conserve la capacité de s'agréger en tétramère et de lier l'inducteur, mais il ne peut pas se lier à l'opérateur. Le fragment N-terminal contenant les résidus 1-59 est décrit comme étant « la longue pièce maîtresse ». Elle peut être encore clivée par la trypsine à la position 51, engendrant un fragment comportant les résidus 1 à 51 qui est appelé « la petite pièce maîtresse ».

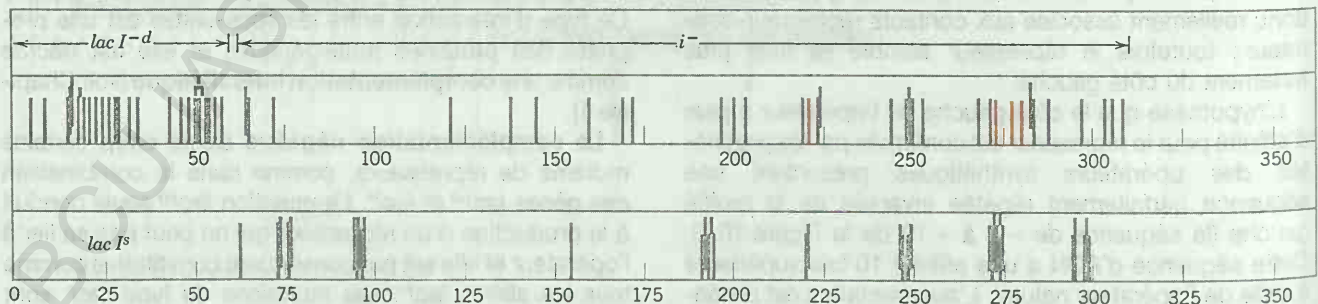
La petite et la longue pièces maîtresses conservent leur capacité à se lier à l'ADN et, en présence d'opérateurs, elles possèdent le même profil de contact que celui obtenu avec le répresseur intact (bien que leur affinité de liaison soit diminuée). La capacité des pièces maîtresses à se lier à l'opérateur suggère que leur structure est indépendante du reste de la molécule.

Ceci est en accord avec le modèle dans lequel le site

Figure 10.8

Les mutations dans le gène $lacI$ permettent d'identifier des domaines pour différentes fonctions.

La distance le long du gène est numérotée en codons. La carte supérieure montre que les mutations $lacI^-$ se produisent tout au long du gène, mais que les mutations $lacI^d$ se produisent seulement dans l'extrémité N-terminale. Le rouge indique les mutations qui sont $lacI^-$ où les monomères ne peuvent s'agréger en tétramères. La carte inférieure montre que les mutations $lacI^s$ ne commencent qu'à partir du résidu 62, et se groupent grossièrement entre les positions 150 et 300, généralement séparées par 60 résidus ou plus.



de fixation à l'ADN réside dans les 50 acides aminés N-terminaux constituant un **bras** ou une **protubérance** sur le corps de la protéine. Le bras peut être connecté au cœur par une région **charnière** qui s'étend des acides aminés 50 à 80. Le reste de la protéine, les acides aminés 81 à 360 sont responsables de l'agrégation en une structure tétramérique et de l'interaction avec l'inducteur.

Si les pièces maîtresses sont vraiment des protubérances ayant la capacité de se lier à l'ADN, quelle est alors la fonction de la structure tétramérique et comment entre-t-elle en contact avec l'ADN ? Un modèle suggère que le répresseur possède la forme d'une cloche allongée ou d'un cylindre de 115 Å de long dont l'axe peut s'étendre à travers l'ADN avec un angle superficiel de 15-20°. Le cœur des 4 sous-unités est en contact dans la région centrale ; les sites de liaison des pièces maîtresses s'étendent aux extrémités des molécules, ordonnées en paires. Il est possible que seulement deux des quatre sous-unités soient en contact avec l'ADN, les deux autres sites de liaison restant inoccupés.

Bien que le cœur de la sous-unité ne reconnaisse pas l'opérateur, il se lie à l'ADN d'une manière non spécifique, si bien qu'il est possible que lui aussi réalise des contacts (bien qu'ils ne soient pas suffisants pour influencer le profil de protection de l'ADN par rapport à ce qui est obtenu avec la pièce maîtresse). Le modèle permet de postuler que le complexe répresseur-ADN est une protéine tétramérique disposée symétriquement sur l'ADN, avec seulement une petite proportion (20 %) de la chaîne polypeptidique intimement associée à l'ADN.

QUITTER L'ADN ET STOCKER L'EXCÉDENT DE RÉPRESSEUR

Un tétramère de répresseur est fortement lié à l'opérateur. Un inducteur arrive et se lie au répresseur. Comment la protéine quitte-t-elle l'ADN ?

Certains inducteurs entraînent des réductions caractéristiques de l'affinité du répresseur pour l'opérateur *in vitro*. Ces changements sont corrélés à l'efficacité des inducteurs *in vivo*. Ceci signifie que le mécanisme d'induction est associé à une diminution de l'interaction entre l'opérateur et le répresseur. Comment ceci s'accomplit-il ? La vitesse avec laquelle le répresseur se dissocie de l'opérateur est plutôt lente, avec un temps de demi-vie *in vitro* de 10-20 minutes. Mais lorsque l'IPTG est ajouté il y a une réduction immédiate de la stabilité du complexe, mesurée par une diminution considérable de son temps de demi-vie.

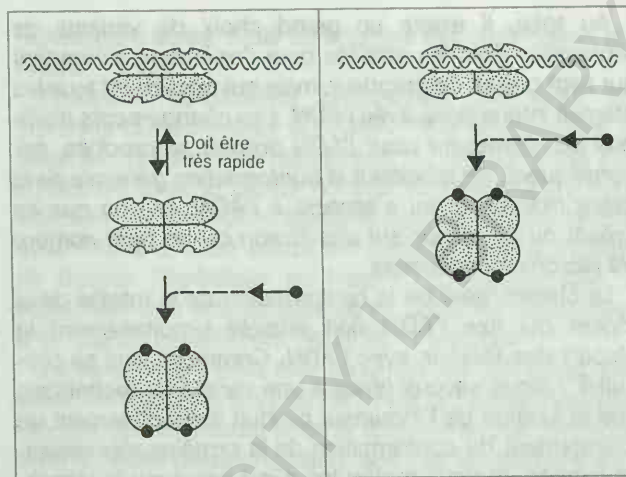


Figure 10.9

L'inducteur se lie-t-il au répresseur libre pour déplacer l'équilibre (à gauche) ou directement au répresseur associé à l'opérateur (à droite) ?

Ce résultat permet la distinction entre les deux modèles d'action du répresseur illustrés dans la Figure 10.9.

- Le modèle d'équilibre (à gauche) fait appel à la notion d'équilibre rapide entre le complexe répresseur-ADN et le répresseur libre ; l'inducteur devrait s'associer à la forme libre du répresseur, ce qui déplace l'équilibre en empêchant la réassociation avec l'ADN.
- Toutefois la vitesse de dissociation du répresseur de l'opérateur (mesurée en l'absence d'inducteur) est bien trop lente pour être compatible avec ce modèle. Ce qui veut dire que *l'inducteur doit se lier directement au répresseur complexé à l'opérateur*. Comme c'est indiqué dans le modèle de droite, la fixation de l'inducteur doit produire un changement dans le répresseur qui le conduit à se dissocier de l'opérateur.

L'interaction du complexe répresseur-IPTG avec l'opérateur peut être étudiée en utilisant de plus fortes concentrations de protéine dans les expériences de protection et d'augmentation de méthylation. Une forte concentration compense la faible affinité du complexe répresseur-IPTG pour l'opérateur. Le complexe produit exactement le même profil de contact avec l'ADN que celui du répresseur libre ou des pièces maîtresses. Un résultat analogue est obtenu avec des mutants de répresseur dont l'affinité pour l'opérateur est augmentée ; ils présentent les mêmes points de contact.

Au total, il existe un grand choix de variants de répresseur dont les affinités pour l'opérateur s'étendent sur sept ordres de grandeur, mais qui possèdent tous les mêmes interactions avec l'ADN. *Les changements d'affinité du répresseur pour l'ADN doivent se produire, par conséquent, en modifiant la conformation générale de la pièce maîtresse qui s'associe à l'ADN, et non pas en créant ou en dissociant une liaison ou un petit nombre de liaisons individuelles.*

Le changement de la conformation de la totalité de la région qui fixe l'ADN doit affaiblir simultanément la plupart des liaisons avec l'ADN. Comment ceci se produit-il ? Nous savons grâce à une variété de techniques que la fixation de l'inducteur produit immédiatement un changement de conformation de la protéine répresseur. La fixation de deux molécules d'inducteur sur le répresseur tétramérique est suffisante pour empêcher la répression. Cependant nous ne savons pas encore comment le changement global de conformation modifie l'affinité du répresseur pour l'opérateur.

Le modèle le plus vraisemblable est que le changement de conformation est transmis du cœur à la pièce maîtresse par l'intermédiaire de la charnière. La **Figure 10.10** présente cette hypothèse sous forme d'un diagramme. La conformation du site de liaison à l'ADN passe, d'un état dans lequel il s'ajuste exactement à la séquence d'ADN, à un état où la pièce maîtresse se trouve dans une conformation un peu différente, qui n'interagit plus aussi fortement avec l'ADN.

Probablement toutes les protéines qui possèdent une forte affinité pour une séquence spécifique, possèdent également une faible affinité pour n'importe quelle séquence d'ADN prise au hasard dans le génome. Un

grand nombre de sites à faible affinité rentreront donc en compétition, avec le petit nombre de sites à forte affinité. Il n'y a qu'un seul site à forte affinité dans le génome d'*E. coli*: l'opérateur. Le restant de l'ADN peut être considéré comme fournissant des sites à faible affinité.

Comment le répresseur est-il partagé entre l'opérateur et le reste de l'ADN ? Qu'arrive-t-il au répresseur lorsqu'il interagit avec l'inducteur et se dissocie de l'opérateur ?

Pour atteindre une spécificité pour l'opérateur dans le contexte du génome, l'affinité d'interaction du répresseur pour l'opérateur doit être beaucoup plus grande que son affinité pour des séquences au hasard. En fait le répresseur fixe 4×10^6 fois mieux l'ADN de l'opérateur que n'importe quelle séquence d'ADN au hasard, de la même longueur (ceci est le rapport de la constante d'association du répresseur pour la séquence *lacO* contre un autre ADN). L'opérateur représente ainsi 26 paires de bases dans un génome de $4,2 \times 10^6$ pb ; l'excès de l'ADN non opérateur sur l'ADN opérateur est donc de 2×10^5 ; autrement dit, il existe 2×10^5 sites de fixation à faible affinité pour le répresseur comparativement au site unique de fixation à forte affinité.

La préférence relative du répresseur pour l'ADN opérateur à l'intérieur du génome bactérien est donnée par le rapport du site de forte affinité (4×10^6) au nombre de sites à faible affinité (2×10^5). Le résultat indique que le répresseur a approximativement 20 fois plus de probabilités de se lier à l'opérateur qu'à n'importe quel autre site.

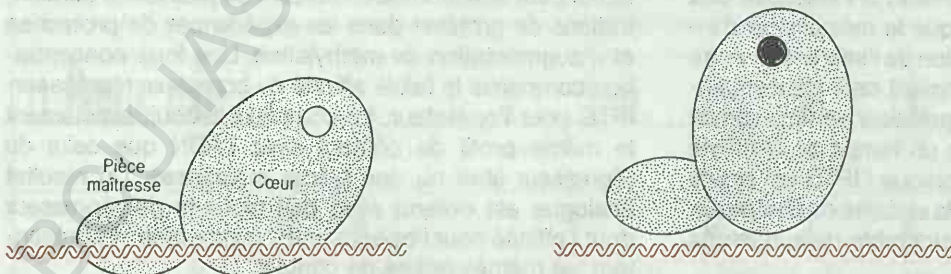
Ceci explique pourquoi des mutations qui réduisent l'affinité de l'opérateur pour le répresseur, ne serait-ce que de 20 à 30 fois, ont un effet suffisant pour être constitutives. Les mutants d'opérateur peuvent être

Figure 10.10

Le déplacement du répresseur implique un changement général de la structure qui affaiblit tous les contacts avec l'opérateur.

A gauche : La pièce maîtresse est liée spécifiquement à l'opérateur, tout au long de l'ADN, pour établir des contacts avec la double hélice.

A droite : L'inducteur change de conformation (illustré par le changement d'angle entre la pièce maîtresse et le cœur), de sorte que la pièce maîtresse ne peut plus entrer en contact avec l'ADN.



débordés par le nombre des sites au hasard. L'affinité spécifique des sites à forte affinité pour les répresseurs n'est plus suffisamment forte pour compenser leur petit nombre.

La conséquence de ces affinités est que dans une cellule non induite, un tétramère de répresseur est toujours lié à l'opérateur. Tous ou la plupart des tétramères restants sont liés au hasard à d'autres régions de l'ADN comme c'est illustré dans la **Figure 10.11**. Il doit y avoir certainement très peu ou pas du tout de tétramères de répresseur libre dans la cellule.

Lorsque l'inducteur se lie au répresseur, l'affinité pour l'opérateur est réduite d'un ordre de grandeur de 10^3 . L'affinité pour d'autres séquences d'ADN reste inchangée. Ainsi dans une cellule induite, probablement tous ou presque tous les tétramères de répresseur seront conservés dans des sites d'ADN au hasard. L'effet de l'induction n'est pas par conséquent d'engendrer des

répresseurs libres, mais de changer leur distribution sur l'ADN.

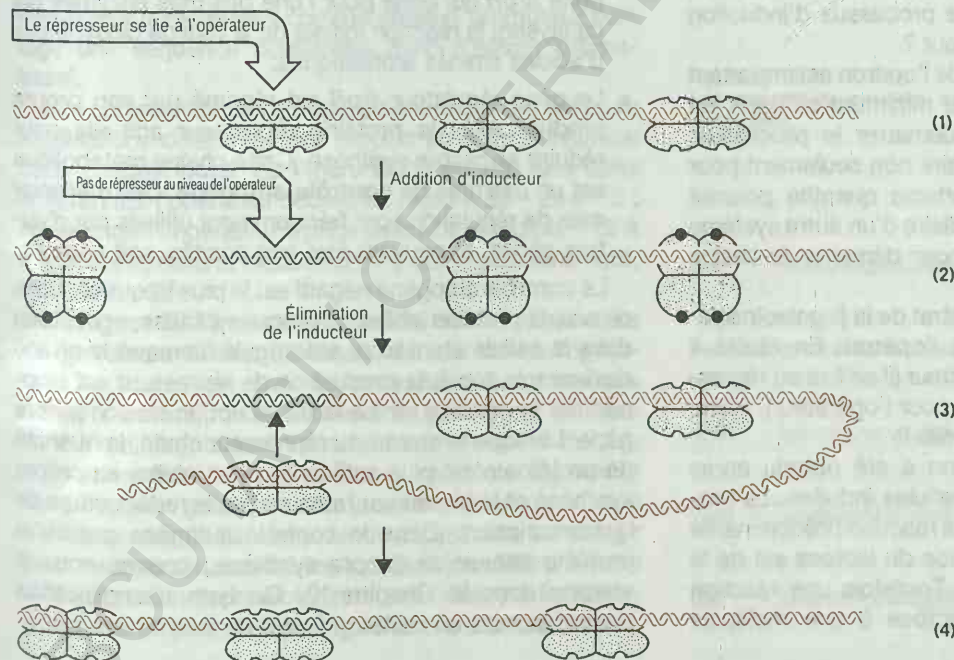
Plusieurs conclusions biologiques découlent du stockage du répresseur sur l'ADN. Plus précisément, la capacité de fixer très rapidement l'opérateur n'est pas en accord avec le temps qui doit être nécessaire pour effectuer de multiples dissociations et réassociations avec l'ADN. L'écart exclut les mécanismes permettant de trouver l'opérateur au hasard, et suggère que le répresseur doit être capable de se déplacer directement d'un site au hasard vers l'ADN de l'opérateur.

Le mouvement peut être accompli soit par un glissement très rapide le long de l'ADN (ce qui paraît improbable, car des obstacles sous la forme d'autres protéines liées à l'ADN doivent exister), soit par un déplacement direct d'un site à un autre site. Ceci est un mécanisme équivalent à celui que nous avons considéré pour l'ARN polymérase (voir Chapitre 9). La réaction doit être facilitée

Figure 10.11

Virtuellement tous les répresseurs de la cellule sont fixés à l'ADN.

- (1) Un tétramère est présent au niveau de l'opérateur.
- (2) Lorsque le tétramère est induit à se déplacer, il se lie à un autre site de l'ADN au hasard. Les autres tétramères interagissent également avec l'inducteur mais ceci n'affecte pas leur affinité générale pour l'ADN.
- (3) Le répresseur peut se déplacer d'un site à un autre dans l'ADN par simple déplacement.
- (4) Lorsque l'inducteur est retiré du milieu, l'opérateur interagit à nouveau avec un tétramère du répresseur qui a quitté un site au hasard pour restaurer l'état non induit.



par la présence des quatre sites de fixation par tétramère alors qu'à chaque fois deux seulement sont en contact avec l'ADN.

La possibilité d'un grand nombre de sites à faible affinité d'entrer en concurrence avec le petit nombre de sites à forte affinité implique que l'interaction du répresseur avec l'opérateur est très sensible à la concentration totale d'ADN et à la concentration totale du répresseur dans la cellule. La différence d'expression de l'opéron entre l'état induit et l'état réprimé est d'un facteur 10^3 . En d'autres termes, même lorsque l'inducteur est absent, il existe une expression basale de 0,1 % du niveau induit. Il pourrait être réduit s'il y avait plus de protéine répresseur, et inversement augmenté s'il y en avait moins. Ainsi paraît-il impossible d'établir une répression même faible, s'il y a moins de 10 molécules de répresseur par cellule, et de même il paraît difficile d'induire l'opéron s'il y en a beaucoup trop.

UN PARADOXE D'INDUCTION

L'opéron lactose possède le gène structural (*lacZ*) qui code pour l'enzyme β -galactosidase, son activité est requise pour métaboliser le sucre ; l'opéron inclut aussi le gène (*lacY*) qui code pour la protéine impliquée dans le transport du substrat à l'intérieur de la cellule. Toutefois si l'opéron est réprimé, comment l'inducteur entre-t-il dans la cellule pour démarrer le processus d'induction en se liant à la protéine répresseur ?

Le niveau basal d'expression de l'opéron est important pour assurer toujours un niveau minimum suffisant de protéine dans la cellule pour démarrer le processus. L'expression basale est nécessaire non seulement pour l'entrée de l'inducteur (une certaine quantité pourrait entrer, par ailleurs, par l'intermédiaire d'un autre système de transport) mais également pour démarrer la chaîne métabolique.

Bien que le lactose soit le substrat de la β -galactosidase, il n'est pas l'inducteur de l'opéron. En réalité il possède un léger effet anti-inducteur (il se fixe au répresseur en augmentant son affinité pour l'opérateur). Alors, comment l'induction se produit-elle ?

L'inducteur naturel de l'opéron a été obtenu après purification du répresseur des cellules induites. La protéine était fixée à l'**allolactose**. La réaction prédominante de la β -galactosidase en présence du lactose est de le cliver en glucose et galactose. Toutefois une réaction « secondaire » transfère le galactose à une molécule

acceptrice, produisant (principalement) l'allolactose et le galactobiose. Puisque l'accepteur utilisé pour produire l'allolactose est en fait le glucose, la réaction aboutit à un réarrangement moléculaire.

Le schéma complexe de l'induction *in vivo* implique l'entrée d'un peu de lactose, une partie est transformée en allolactose par l'enzyme. L'allolactose induit ensuite l'opéron. La forte induction par une petite quantité d'allolactose, qui continue à être synthétisée, est suffisante pour contrebalancer le faible effet anti-inducteur du substrat lactose.

LA RÉPRESSION PEUT SE PRODUIRE A DE MULTIPLES LOCI

Le répresseur *lactose* agit uniquement sur l'opérateur du groupe de gènes *lac Z, Y et A*. Cependant d'autres répresseurs peuvent interagir avec plusieurs opérateurs. Un exemple est représenté par le répresseur tryptophane *trp* qui régule trois groupes de gènes indépendants :

- L'opérateur du groupe de gènes structuraux *trp* *EDBCA* qui contrôle d'une manière coordonnée la synthèse des enzymes qui produisent le *trp* à partir de l'acide chorismique.
- L'opérateur d'un autre locus contrôlant l'expression du gène *aroH* qui code pour l'une des trois enzymes qui catalysent la réaction initiale de la voie de biosynthèse d'acides aminés aromatiques.
- Le gène régulateur *trpR* est réprimé par son propre produit. Ainsi la protéine répresseur agit-elle pour réduire sa propre synthèse. Cette chaîne métabolique est un exemple de **contrôle autogène**. Ces mécanismes de régulation sont fréquemment utilisés par d'autres gènes, et peuvent être soit positifs, soit négatifs.

Le contrôle autogène négatif est le plus fréquent ; dans ce cas, la protéine inhibe sa propre synthèse, son niveau dans la cellule étant ainsi autorégulé. Lorsque le niveau devient très élevé, la production de répresseur est stoppée car la protéine inhibe la transcription de son propre gène. Lorsque le niveau du répresseur chute, la quantité de protéine n'est plus suffisante pour inhiber sa propre synthèse et le niveau est restauré par le redémarrage de la transcription. (Dans le contrôle autogène positif, la protéine stimule sa propre synthèse ; comme nous le verrons dans le Chapitre 12. Ce type d'amplification permet la mise en route ou l'arrêt de la synthèse.)

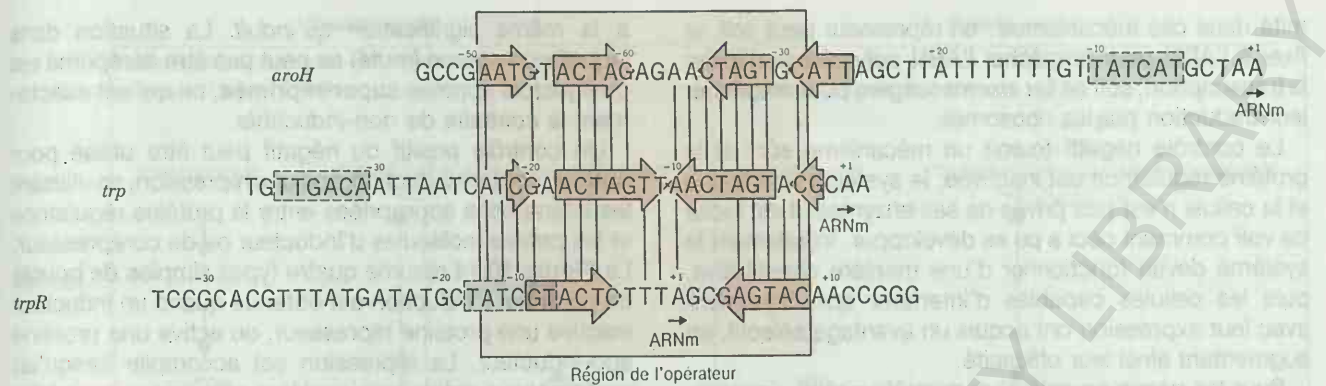


Figure 10.12
Le répresseur *trp* reconnaît les opérateurs au niveau de trois loci différents.

Les lignes verticales représentent les bases conservées dans les opérateurs. Les flèches indiquent les régions de symétrie « dyad ».

Une séquence homologue à l'opérateur qui s'étend sur 21 pb est présente dans les trois loci ; elle est la cible du répresseur *trp*. La conservation de la séquence est indiquée dans la **Figure 10.12**. Chaque opérateur contient des éléments de symétrie « dyad » appréciables (mais non identiques). Vraisemblablement les caractéristiques communes des trois opérateurs incluent les points de contacts importants avec le répresseur *trp*. Ceci explique comment une protéine répresseur peut agir sur différents loci : *chaque locus possède une copie de la séquence d'ADN spécifique de liaison reconnue par le répresseur* (exactement comme chaque promoteur partage une séquence consensus avec d'autres promoteurs).

Une caractéristique des opérateurs dispersés est leurs localisations différentes par rapport au point d'initiation dans chaque locus. La **Figure 10.12** montre que dans l'opéron *trp*, l'opérateur couvre les régions de -23 à -3 ; en revanche l'opérateur *trpR* s'étend de la région -12 à +9, et enfin, dans le locus *aroH*, il s'étend plus en amont, de -49 à -29.

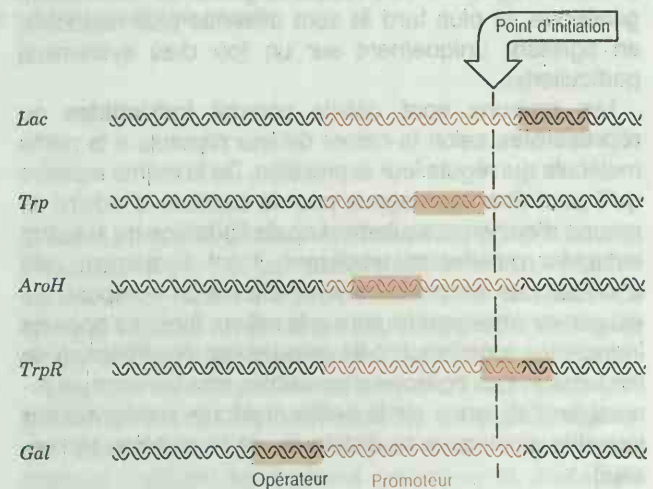
La **Figure 10.13** résume la variété des relations entre les opérateurs et les promoteurs, dans le cas où l'opérateur se trouve en aval du promoteur (comme dans l'opéron lactose), ou à l'intérieur du promoteur (comme dans les différents loci répondant au répresseur *trp*), ou enfin immédiatement en amont du promoteur (comme dans l'opéron galactose, où la nature de la répression n'est pas claire). La capacité du répresseur à interagir avec les opérateurs dont les localisations diffèrent dans différents promoteurs cibles renforce l'idée que *la répression est un problème de blocage d'accès au promoteur*.

LA DISTINCTION ENTRE UN CONTRÔLE POSITIF ET NÉGATIF

Un opéron répond à l'absence de protéine régulatrice par des contrôles positifs ou négatifs. Les caractéristiques de ces deux types de systèmes de contrôle sont des images en miroir.

Les gènes qui sont contrôlés négativement s'expriment *jusqu'à ce qu'ils soient réprimés par la protéine répresseur*. Toute action interférant avec l'expression génétique peut fournir un contrôle négatif, mais il existe une unifor-

Figure 10.13
Les opérateurs peuvent se trouver à différentes positions du promoteur.



mité dans ces mécanismes : un répresseur peut soit se fixer à l'ADN pour empêcher l'ARN polymérase d'initier la transcription, soit se lier aux messagers pour empêcher leur traduction par les ribosomes.

Le contrôle négatif fournit un mécanisme sûr : si la protéine régulatrice est inactivée, le système fonctionne et la cellule n'est pas privée de ses enzymes. Il est facile de voir comment ceci a pu se développer. Initialement le système devait fonctionner d'une manière constitutive, puis les cellules capables d'interférer spécifiquement avec leur expression ont acquis un avantage sélectif, en augmentant ainsi leur efficacité.

Pour les gènes soumis à un contrôle positif, l'expression est possible *uniquement lorsqu'une protéine régulatrice active est présente*. Le mécanisme de contrôle de l'opéron est exactement l'inverse du contrôle négatif : la protéine régulatrice, au lieu de l'empêcher, est indispensable à l'initiation de la transcription. Cette protéine interagit avec l'ADN et l'ARN polymérase pour activer l'étape d'initiation. La protéine est appelée un **apo-inducteur**. D'autres contrôles positifs fonctionnent grâce à la substitution globale des facteurs sigma préexistants par d'autres ; ces facteurs modifient la sélection des promoteurs (Chapitre 9) ; ou bien grâce aux facteurs anti-terminateurs, qui modifient la reconnaissance des sites de terminaison (Chapitre 11).

Il est beaucoup plus difficile de savoir comment un contrôle positif s'est développé, car la cellule devait posséder la capacité d'exprimer ces gènes avant même que le contrôle existe. Vraisemblablement, certains des éléments du système régulateur ont changé de rôle. Probablement, au début ils ont été utilisés comme des éléments régulateurs de l'appareil global de l'expression génétique, et plus tard ils sont devenus plus restrictifs, en agissant uniquement sur un (ou des) système(s) particulier(s).

Les opérons sont définis comme **inductibles** ou **répressibles** selon la nature de leur réponse à la petite molécule qui régule leur expression. De la même manière qu'il peut être avantageux pour la bactérie d'induire un groupe d'enzymes seulement après l'addition du substrat inducteur qu'elles métabolisent, il est également utile d'inhiber des enzymes qui synthétisent un composé qui est convenablement fourni par le milieu. Ainsi les opérons inductibles fonctionnent-ils uniquement en présence de l'inducteur. Les opérons répressibles fonctionnent seulement en l'absence de la petite molécule **corépresseur** (appelée ainsi, pour la distinguer de la protéine répresseur).

La terminologie utilisée pour les systèmes répressibles décrit l'état actif de l'opéron comme **déréprimé**, ce qui

a la même signification qu'*induit*. La situation dans laquelle un opéron (muté) ne peut pas être déréprimé est quelquefois appelée **super-réprimée**, ce qui est exactement le contraire de *non-inductible*.

Un contrôle positif ou négatif peut être utilisé pour obtenir, soit une induction, soit une répression, en utilisant les interactions appropriées entre la protéine régulatrice et les petites molécules d'inducteur ou de corépresseur. La **Figure 10.14** résume quatre types simples de boucle de contrôle. L'induction est obtenue quand un inducteur inactive une protéine répresseur, ou active une protéine apo-inducteur. La répression est accomplie lorsqu'un corépresseur active une protéine répresseur ou inhibe une protéine apo-inducteur.

Les conséquences génétiques de l'inactivation de la protéine régulatrice peuvent être utilisées pour distinguer les systèmes à contrôle positif ou négatif. L'inactivation de la protéine répresseur entraîne une constitutivité récessive (ou dérépession), typique des systèmes à contrôle négatif. D'autre part une mutation qui inactive un apo-inducteur entraîne une non-inductibilité récessive, ou super-répression, qui caractérise les systèmes à contrôle positif.

Le tryptophane est le corépresseur qui active la protéine répresseur. Ceci représente un mécanisme classique de répression ; un exemple en est présenté dans la **Figure 10.14** et il est montré plus en détail dans la **Figure 10.15**. Dans des conditions où l'approvisionnement en tryptophane est maximum, l'opéron est réprimé car le complexe répresseur co-répresseur est lié à l'opérateur. Lorsque le tryptophane manque, le répresseur est inactif ; son affinité pour l'opérateur est réduite, et il se trouve stocké ailleurs sur l'ADN.

Le répresseur *tryptophane* agit sous forme d'un tétramère de quatre sous-unités identiques de 12 500 daltons chacune. Il existe environ 20 molécules de tétramère par bactérie. Le répresseur se lie à l'opérateur *in vitro* uniquement en présence de tryptophane. La seule différence significative avec l'opéron *lactose* est que le tryptophane *augmente* l'affinité du répresseur pour l'opérateur.

LA RÉPRESSION CATABOLIQUE IMPLIQUE UNE RÉGULATION POSITIVE AU NIVEAU DU PROMOTEUR

Jusqu'à présent nous avons défini le promoteur comme une séquence d'ADN capable de lier l'ARN polymérase, pour permettre l'initiation de la transcription. Toutefois il existe des promoteurs au niveau desquels l'ARN polymérase ne peut initier la transcription sans l'aide d'une protéine auxiliaire. De telles protéines constituent des

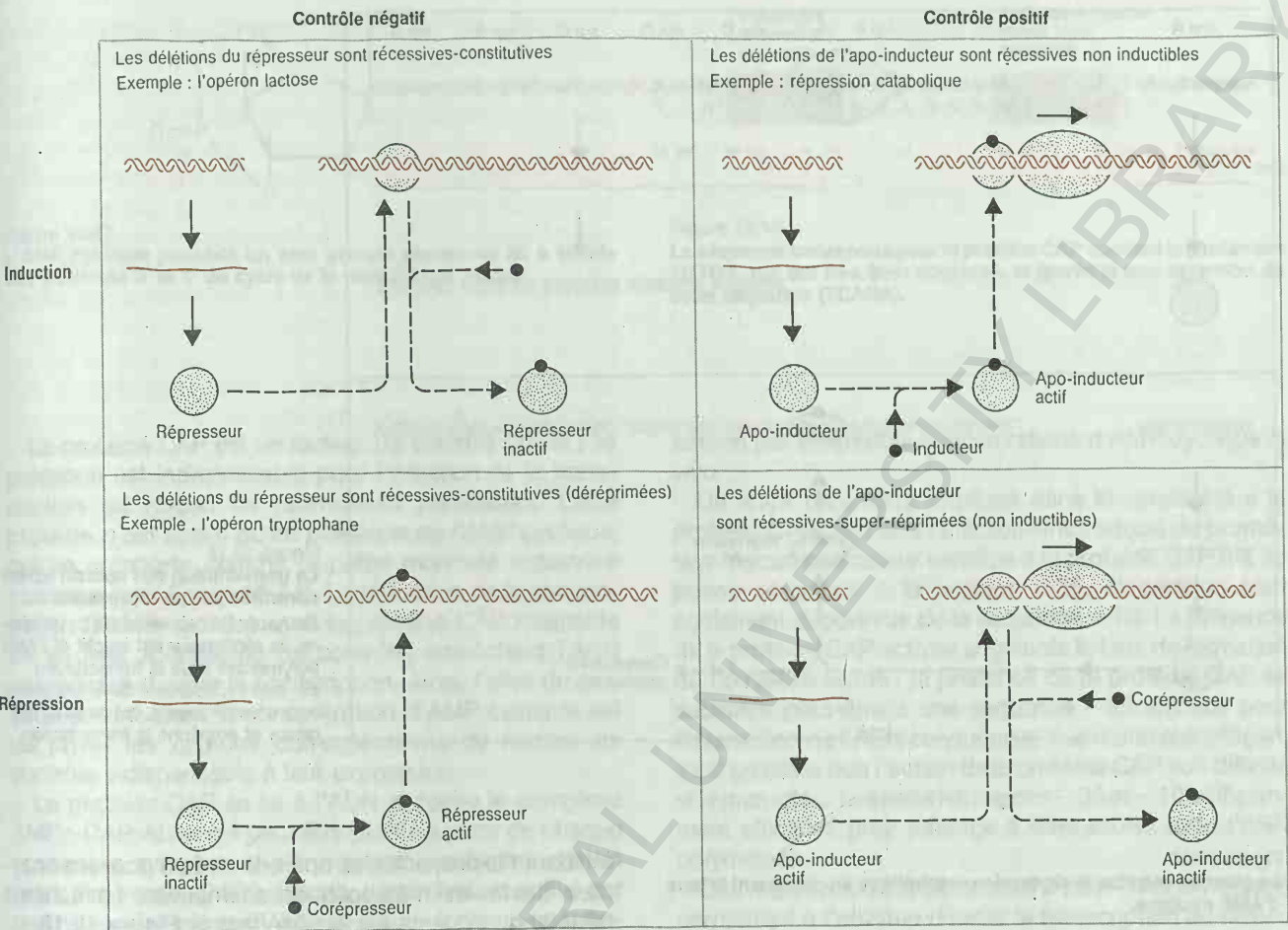


Figure 10.14 Les circuits de contrôle sont très souples et peuvent permettre un contrôle positif ou négatif, c'est-à-dire une induction ou une répression.

régulateurs positifs, car leur présence est nécessaire pour démarrer la transcription. Plusieurs régulateurs positifs sont connus ; certains sont codés par des phages, et d'autres sont présents dans la cellule hôte. Pour commencer il est nécessaire de dire que nous n'avons pas exactement compris la différence qui existe entre les promoteurs qui fonctionnent seuls et ceux qui nécessitent un régulateur positif.

Un des régulateurs positifs les plus répandus est la protéine qui contrôle l'activité d'un grand groupe d'opérons répondant tous aux conditions nutritionnelles chez

E. coli. Lorsque le glucose est disponible comme source d'énergie, il est utilisé préférentiellement à d'autres sucres. Ainsi quand *E. coli* se trouve dans un milieu qui contient à la fois du glucose et du lactose (par exemple, elle métabolise le glucose et inhibe l'utilisation du lactose).

Le choix est obtenu en empêchant l'expression des gènes de l'opéron *lactose* ; cet effet est appelé **répression catabolique**. Le même effet est trouvé dans d'autres opérons, incluant les opérons *galactose* et *arabinose*. Ainsi la répression catabolique représente-t-elle un système de coordination général qui privilégie le glucose, en

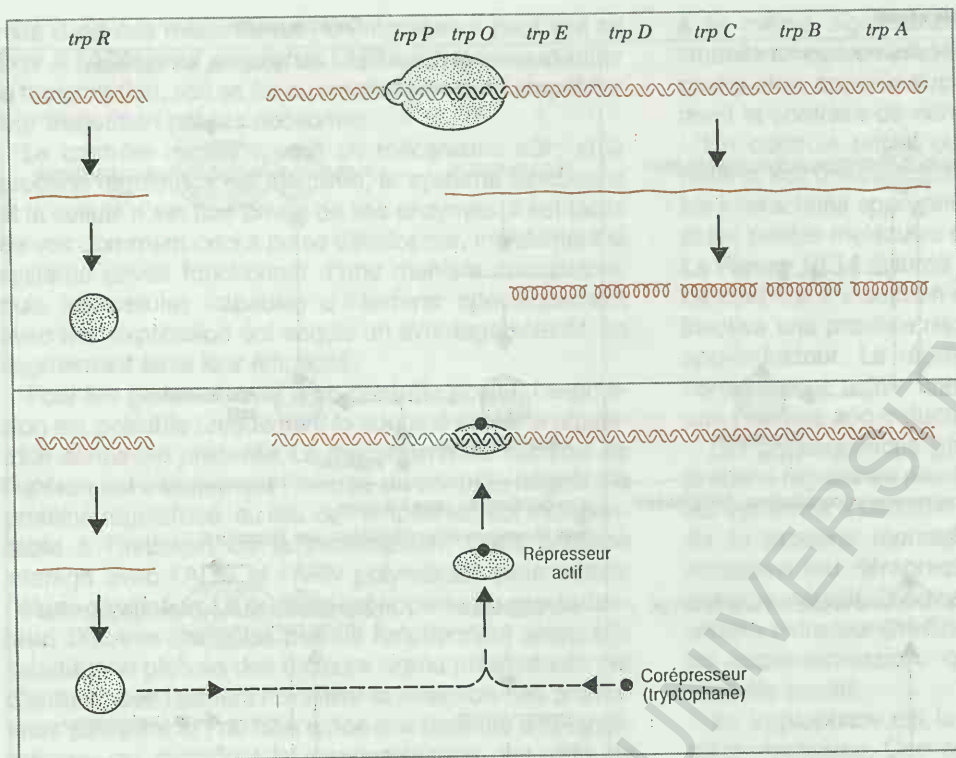
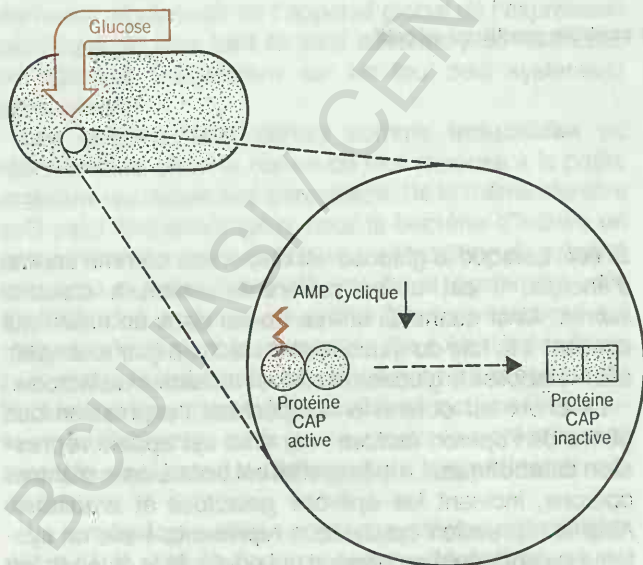


Figure 10.15
La transcription de l'opéron *trp* est contrôlée par un répresseur
En haut : En l'absence de tryptophane, le répresseur est inactif et l'ARN polymérase initie la transcription.
En bas : L'addition de tryptophane active le répresseur, qui se lie à l'opérateur et empêche la transcription.

Figure 10.16
Le glucose entraîne la répression catabolique en diminuant le taux d'AMP cyclique.



inhibant l'expression des opérons codant pour les enzymes des voies métaboliques alternatives. Le schéma général du système est illustré dans la **Figure 10.16**.

La répression catabolique repose sur la capacité du glucose à réduire le niveau de l'AMPc dans la cellule. Nous ne savons pas exactement comment cette inhibition est obtenue, peut-être en réduisant la vitesse de synthèse de l'AMPc. Quoi qu'il en soit, l'expression des opérons contrôlés par la répression catabolique montre une relation inverse avec la concentration d'AMPc.

L'AMP cyclique est synthétisé par l'enzyme **adénylatecyclase**. La réaction utilise l'ATP comme substrat, en introduisant dans la molécule un pont formé par une liaison phosphodiester en 3'-5'. La **Figure 10.17** montre la structure de l'AMPc. Les mutants du gène codant pour l'adénylate cyclase (*cya*⁻) ne présentent pas de répression catabolique.

Des mutations alternatives connues comme *cap*⁻ ou *crp*⁻ permettent d'identifier la protéine régulatrice qui agit directement sur les opérons cibles. La protéine est connue sous le nom de protéine CAP (catabolic activator protein, protéine d'activation catabolique) ou CRP (cyclic AMP receptor protein, protéine acceptrice de l'AMPc).

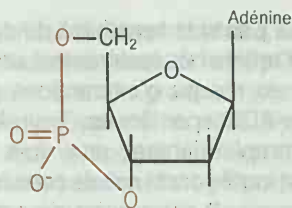


Figure 10.17
L'AMP cyclique possède un seul groupe phosphate lié à la fois aux positions 3' et 5' du cycle de la molécule de ribose.

La protéine CAP est un facteur de contrôle positif ; sa présence est indispensable pour l'initiation de la transcription au niveau de promoteurs particuliers. Cette protéine *n'est active qu'en présence de l'AMP cyclique*, qui se comporte comme la petite molécule inductrice classique (voir Figure 10.14). La réduction de la concentration d'AMP cyclique rend la protéine CAP incapable de se fixer à la région de contrôle, empêchant l'ARN polymérase d'initier la transcription. Ainsi, l'effet du glucose en réduisant la concentration d'AMP cyclique est de priver les opérons correspondants du facteur de contrôle indispensable à leur expression.

La protéine CAP se lie à l'ADN et forme le complexe AMPc-CAP-ADN, qui peut être purifié à partir de chaque promoteur avec lequel elle interagit. Le facteur est un homodimère de 22 500 daltons. Le dimère est activé par une molécule unique d'AMP cyclique. A chaque promoteur, le site de liaison CAP de 22 pb est probablement lié à un seul dimère. Les sites de fixation constituent des variations de la séquence consensus qui est montrée dans la Figure 10.18.

Les mutations empêchant l'action de la protéine CAP sont souvent localisées à l'intérieur de la courte séquence fortement conservée ^{TGTGA} _{ACACT}. Cette très courte séquence semble être l'élément essentiel de la reconnaissance. Elle est liée à l'une des sous-unités du dimère.

Que reconnaît la seconde sous-unité du dimère ? Certains sites de fixation contiennent une séquence inversée répétée de la séquence consensus ; dans ce cas, cette séquence pourrait être le site de liaison de la seconde sous-unité. Cependant, la plupart des sites ne sont pas symétriques et, à leur niveau, la seconde sous-unité doit se fixer à des séquences différentes.

La protéine CAP se lie très fortement aux sites qui contiennent deux copies inversées du pentamère. La classification par affinité de liaison pour la protéine CAP permettrait d'expliquer pourquoi les différents gènes sont

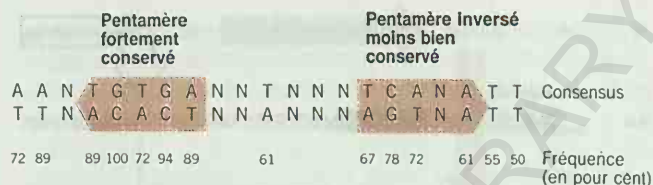


Figure 10.18
La séquence consensus pour la protéine CAP contient le pentamère TGTGA, qui est très bien conservé, et (parfois) une inversion de cette séquence (TCANA).

activés par différentes concentrations d'AMP cyclique *in vivo*.

Un autre paramètre impliqué dans la sensibilité à la protéine CAP peut être l'efficacité intrinsèque du promoteur. Aucun promoteur sensible à la protéine CAP n'a de bonne séquence à la position -35, et certains sont également dépourvus de la séquence -10. La présence de la protéine CAP activée augmente le taux de formation du complexe fermé ; la présence de la protéine CAP se substitue peut-être à une séquence -35 efficace pour l'interaction de l'ARN polymérase. Vue d'une autre façon, il est possible que l'action de la protéine CAP soit difficile si le promoteur possède les régions -35 et -10 suffisamment efficaces pour interagir à elles seules avec l'ARN polymérase.

Des mutations de la sous-unité β de l'ARN polymérase permettent à l'enzyme d'initier la transcription au niveau des promoteurs dépendants de la protéine CAP en l'absence de protéine. Ainsi le rôle de la protéine CAP est peut-être d'aider la polymérase à franchir une étape de l'initiation.

L'action de la protéine CAP est énigmatique car ses sites de liaison se trouvent toujours à des localisations variables du site d'initiation des différents opérons qu'elle contrôle. De plus le pentamère TGTGA peut se trouver dans n'importe quelle orientation. Les trois exemples résumés dans la Figure 10.19 recouvrent l'étendue des localisations.

Le site de liaison de la protéine CAP peut être adjacent au promoteur. Par exemple, dans l'opéron *lactose* la région protégée par la protéine s'étend de la position -72 à la position -52. Il est possible que deux dimères CAP interagissent. Le profil de liaison est en accord avec la présence de la protéine CAP au niveau de la même face de l'ADN que celle liant l'ARN polymérase. Cette localisation permet de placer chacune des protéines à portée l'une de l'autre.

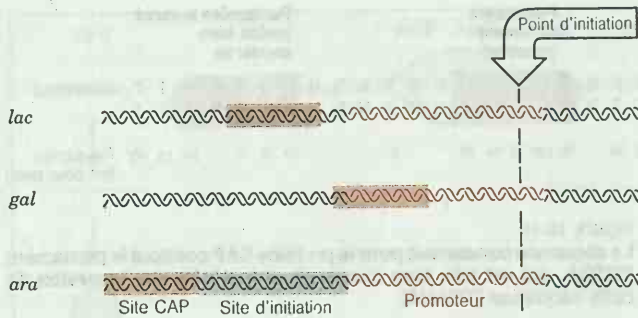


Figure 10.19

La protéine CAP peut se lier à des sites différents par rapport à l'ARN polymérase.

Parfois, le site de fixation de la protéine CAP se trouve à l'intérieur du promoteur. Par exemple, pour le locus *gal*, le site de fixation de la protéine s'étend de la position - 50 à la position - 23. Il semble bien qu'un seul dimère CAP soit fixé, probablement en interaction intime avec l'ARN polymérase, car le site de fixation de la protéine CAP s'étend dans la région généralement protégée par l'ARN polymérase.

Pour les autres opérons, le site de fixation de la protéine CAP se localise bien en amont du promoteur. Dans l'exemple d'*ara*, la région de fixation protégée de la protéine se trouve très éloignée du site d'initiation et s'étend de la position - 107 à - 78. Dans ce cas, la CAP ne peut pas interagir avec l'ARN polymérase, car il y a d'autres protéines régulatrices fixées dans la région entre la protéine CAP et l'ARN polymérase.

La diversité de la localisation des sites de fixation de la protéine CAP rend difficile l'établissement d'un modèle unique de ses actions. Deux types de modèles généraux ont été proposés, reposant sur l'interaction protéine-protéine ou protéine-ADN.

La formation d'un complexe fermé serait aidée par les interactions protéine-protéine lorsque deux protéines (ou plus) sont fixées sur l'ADN. Une légère interaction entre les deux protéines serait suffisante pour augmenter la capacité de fixation de l'ARN polymérase sur le promoteur. La géométrie de l'interaction du complexe CAP-ARN polymérase doit être différente dans les opérons *lac* et *gal*; dans l'opéron *ara*, il est nécessaire qu'il y ait une interaction faisant intervenir trois protéines.

Une autre possibilité est que les effets de la protéine CAP résident entièrement dans son interaction avec l'ADN. Il est possible que, pour les promoteurs sensibles à la protéine CAP, l'ARN polymérase ne puisse pas accomplir la fusion initiale de l'ADN. Si la protéine CAP assume cette fonction, une région de séparation de brin pourrait se propager à partir du site de fixation.

Nous ne savons pas si la sensibilité de divers opérons à la protéine CAP reflète nécessairement un mode univoque de contrôle au niveau d'interactions protéine-protéine ou protéine-ADN spécifiques. Toutefois, le même objectif est accompli : inhiber une voie métabolique lorsqu'elle devient inutile car la cellule possède la quantité suffisante de glucose. Il est à noter de nouveau que le contrôle coordonné, pour la régulation de type négatif ou positif, peut s'étendre à des loci dispersés grâce à la répétition des sites de fixation des protéines régulatrices.

LES PÉRIODES CRITIQUES PROVOQUENT LA RÉPONSE « STRINGENTE »

Lorsque les bactéries se trouvent dans des conditions de croissance si pauvres qu'elles manquent d'apport suffisant en acides aminés pour maintenir la synthèse protéique, elles arrêtent un grand nombre d'activités. Ceci est appelé la **réponse extrême (stringent)**. Nous pouvons la considérer comme un mécanisme de survie dans des conditions difficiles : les bactéries économisent leurs ressources en engageant le minimum d'activités jusqu'à ce que les conditions nutritionnelles soient rétablies ; alors, elles démarrent à nouveau toutes leurs activités métaboliques.

La réponse « stringente » entraîne une réduction importante (10 à 20 fois) de la synthèse des ARNr et des ARNt, qui, à elle seule, suffit à réduire le taux de la synthèse totale d'ARNs à seulement 5 à 10 % de son niveau initial. La synthèse de certains ARN messagers est réduite ; globalement, il y a une réduction de 3 fois de la synthèse globale des ARN messagers. Le taux de dégradation des protéines est augmenté. De nombreux ajustements métaboliques surviennent, tels qu'en témoignent les synthèses réduites de nucléotides, de glucides, de lipides, etc.

La carence en un seul acide aminé, ou une mutation inactivant une aminoacyl-ARNt synthétase, est suffisante pour déclencher la réponse « stringente ». La mise en marche de cette série d'événements est provoquée par la **présence d'un ARNt non chargé dans le site A des ribosomes**. Dans des conditions normales, seuls les aminoacyl-ARNt sont localisés dans le site A par l'EF-Tu (voir le Chapitre 5). Mais quand aucun aminoacyl-ARNt n'est disponible pour s'associer à un codon particulier, les ARNt non chargés deviennent alors capables de s'engager dans le site A du ribosome. De ce fait, ils bloquent toute progression ultérieure des ribosomes, en déclenchant une **réaction stérile**.

Lorsque les cellules sont privées d'acides aminés, elles accumulent deux types de nucléotides inhabituels : le

ppGpp (la guanosine tétraphosphate, où les positions 5' et 3' sont diphosphorylées) et le **pppGpp** (la guanosine pentaphosphate avec un groupe triphosphate à l'extrémité 5' et un diphosphate du côté 3'). Dans toutes les souches d'*E. coli*, la réponse « stringente » est accompagnée d'une accumulation de ppGpp ; le pppGpp peut, quant à lui, être ou non produit. Ces nucléotides sont des petites molécules « effectrices », dont l'activité est de modifier la conformation des protéines auxquelles elles s'associent.

Les composants impliqués dans la production de ppGpp et pppGpp ont été identifiés par des mutations qui suppriment la réponse « stringente ». Pour les **mutants relâchés**, la carence en acides aminés n'entraîne aucune réduction de synthèse d'ARN stable ni aucune des autres réactions composant la réponse « stringente ».

Les mutations relâchées les plus rencontrées se trouvent dans le gène *relA*, qui code pour une protéine appelée le « **facteur stringent** ». Malgré sa faible concentration (moins d'une molécule pour 200 ribosomes), ce facteur est associé aux ribosomes. Ainsi, il est vraisemblable que seule une minorité de ribosomes soit capable de produire la réponse « stringente ».

Les ribosomes isolés à partir des bactéries « stringentes » peuvent synthétiser *in vitro* le ppGpp et le pppGpp pourvu que le site A soit occupé par un ARNt non chargé qui reconnaît spécifiquement le codon. Les ribosomes purifiés des mutants relâchés n'assurent pas cette réaction, mais ils sont capables de l'accomplir si le facteur « stringent » est ajouté.

La Figure 10.20 montre les voies de synthèse des nucléotides à guanines inhabituels. Le facteur « stringent » est une enzyme qui catalyse la réaction de synthèse dans laquelle l'ATP est utilisé pour donner un groupe pyrophosphate au groupe 3'OH du 5' GTP ou GDP. Le GTP est le substrat le plus fréquent. Toutefois, le pppGpp peut être transformé en ppGpp par plusieurs enzymes ; parmi celles qui sont capables de réaliser cette déphosphorylation se trouvent les facteurs de traduction EF-Tu et EF-G. La production de ppGpp via pppGpp est la voie la plus utilisée.

Quels sont les composants impliqués dans la réaction stérile ? Les mutations dans un autre locus entraînant le phénotype relâché se révèlent être situées dans la protéine L11 de la sous-unité 50S, c'est-à-dire dans le voisinage immédiat des sites A et P. La présence d'un

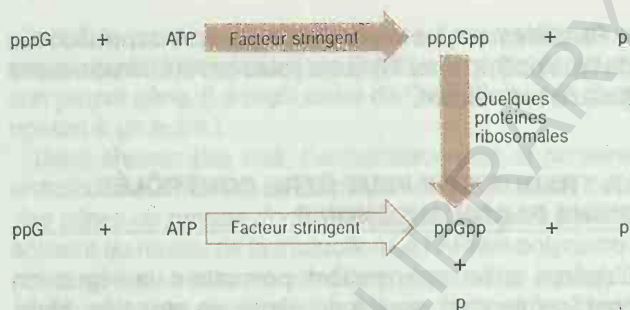


Figure 10.20

Le « facteur stringent » catalyse les réactions de synthèse de pppGpp et ppGpp ; les protéines ribosomales peuvent déphosphoryler les pppGpp pour donner le ppGpp.

appariement correct de l'ARNt non chargé dans le site A pourrait déclencher un changement de conformation des ribosomes ; mais en raison de l'absence d'acide aminé sur l'ARNt, la réaction stérile apparaît à la place du transfert du polypeptide issu du peptidyl-ARNt.

A quoi servent les ppGpp ? Ils pourraient être des effecteurs contrôlant plusieurs réactions, y compris l'inhibition de la transcription. Plusieurs faits ont été rapportés ; parmi eux, deux sont à souligner :

- *L'initiation de la transcription est spécifiquement inhibée dans les promoteurs des opérons codant pour les ARNr.* Les mutations dans les promoteurs régulés dans des conditions « stringentes » peuvent entraîner la perte du contrôle « stringent » ; ceci suggère que la réponse « stringente » nécessite l'interaction avec une séquence spécifique du promoteur.
- *L'étape d'élongation de la transcription de la plupart des matrices est réduite par le ppGpp, en raison de l'augmentation des pauses effectuées par l'enzyme.* Cet effet est responsable de la diminution générale de l'efficacité de la transcription lorsque le ppGpp est ajouté *in vitro*. Nous ne connaissons pas encore la spécificité d'une telle inhibition, mais nous ne serons pas surpris de noter d'importantes variations entre les opérons quant à leur niveau d'inhibition.

Il est intéressant de noter que des nucléotides inhabituels sont utilisés au moins dans deux systèmes de contrôle ayant une fonction de coordination générale. L'ensemble semble être spécifique des bactéries. Bien que l'AMP cyclique soit le second messager classique chez les eucaryotes, il n'est pas certain qu'il soit impliqué dans le contrôle de l'initiation de la transcription *.

* N. d. T. : En fait, très récemment plusieurs promoteurs eucaryotiques spécifiquement contrôlés par l'AMPc ont été décrits. Comme chez les procaryotes, une protéine activée par l'AMPc peut se fixer en un site particulier du promoteur, le CRE (cyclic AMP Response Element, élément de réponse à l'AMPc). Cette protéine a été récemment purifiée. Son activation serait due à sa phosphorylation par une protéine kinase catalysée par l'AMPc plutôt qu'à la fixation de l'AMPc lui-même.

Parallèlement, les recherches intensives de nucléotides de type ppGpp chez les eucaryotes ont été infructueuses jusqu'à maintenant.

LA TRADUCTION PEUT ÊTRE CONTRÔLÉE D'UNE FAÇON AUTONOME

L'opéron a un arrangement permettant la régulation coordonnée d'un groupe de gènes de structure. Mais, superposés à celles-ci, des contrôles supplémentaires peuvent créer des *différences* dans le degré d'expression de chaque gène.

Environ 70 protéines forment l'appareil d'expression génétique. Les protéines ribosomales et les protéines auxiliaires impliquées dans la synthèse protéique représentent les composantes majeures de cet appareil. Les sous-unités de l'ARN polymérase et les autres facteurs de transcription représentent le reste. Une régulation coordonnée garantit la production de ces protéines à un rythme approprié aux conditions de croissance : quand la bactérie croît plus rapidement, elle consacre une plus grande proportion de ses efforts à la production des composants de l'appareil d'expression génétique. Un ensemble de mécanismes est utilisé pour le contrôle de l'expression des gènes codant pour cette machinerie.

Pratiquement chacune de ces protéines est représentée par un gène unique chez *E. coli*. Ces gènes, codant pour les protéines ribosomales, pour les facteurs de synthèse protéique et pour les sous-unités de l'ARN polymérase, sont tous associés et organisés en un petit nombre d'opérons.

L'organisation de six de ces opérons est résumée dans la **Figure 10.21**. Approximativement la moitié des gènes codant pour les protéines ribosomales (souvent appelées **protéines r**) sont répartis dans quatre opérons très proches les uns des autres. Ils sont appelés opérons *str*, *spc*, *S10* et α (chaque dénomination décrit la première fonction qui a été identifiée). Les opérons *rif* et *L 11* sont réunis dans un autre endroit.

Chaque opéron possède un mélange de fonctions. L'opéron *str* contient des gènes codant pour des protéines de la petite sous-unité ribosomale ainsi que pour les facteurs EF-Tu et EF-G. Les opérons *spc* et *S10* contiennent des gènes codant pour des protéines de la petite aussi bien que de la grande sous-unité ribosomales. L'opéron α contient également des gènes codant pour des protéines de deux sous-unités ribosomales, mais aussi pour la sous-unité α de l'ARN polymérase. Le locus *rif* comprend les gènes codant pour des protéines de la grande sous-unité ribosomale et pour les sous-unités β et β' de l'ARN polymérase.

Figure 10.21

Les gènes codant pour les protéines ribosomales, pour les facteurs de synthèse protéique et pour les sous-unités de l'ARN polymérase se trouvent organisés en un petit nombre d'opérons régulés d'une façon autonome.

Les opérons s'étendent sur deux groupes de gènes. L'un contient les opérons *str* — 14.000 pb-*S10*-*spc*- α ; l'autre contient les opérons adjacents *L11*-*rif*. Chaque opéron est écrit de telle sorte que son promoteur soit à l'extrémité gauche. La protéine régulatrice figure à droite et les protéines sujettes à régulation apparaissent sur fond rouge.

Opéron	Les gènes et les protéines (dans l'ordre à partir du promoteur)	Régulateur
<i>str</i>	<i>rpsL-rpsG-fusA-tufA</i> S12 S7 EF-G EF-Tu	S7
<i>spc</i>	<i>rplN-rplX-rplE-rpsN-rpsH-rplF-rplR-rpsE-rplD-rpmO-secY-X</i> L14 L24 L5 S14 S8 L6 L18 S5 L30 L15 Y X	S8
<i>S10</i>	<i>rpsJ-rplC-rplB-rplD-rplW-rplS-rplV-rpsC-rpsQ-rplP-rpmC</i> S10 L3 L2 L4 L23 S19 L22 S3 S17 L16 L29	L4
α	<i>rpsM-rpsK-rpsD-rpoA-rplQ</i> S13 S11 S4 α L17	S4
<i>L11</i>	<i>rplK-rplA</i> L11 L1	L1
<i>rif</i>	<i>rplJ-rplL-rpoB-rpoC</i> L10 L7 β β' L12	L10

Toutes les protéines ribosomales, à l'exception d'une seule, sont nécessaires en concentrations équimolaires, qui doivent être coordonnées avec le niveau des ARNr. Un problème intéressant de l'expression génétique coordonnée se pose lorsque des gènes dont les produits sont nécessaires en quantités équimolaires se trouvent associés à d'autres gènes dont les produits sont nécessaires à des niveaux différents.

La protéine L7/L12 présente une exception ; elle est en effet présente en quatre copies par ribosome. Une autre exception est représentée par le facteur EF-Tu qui est présent à un niveau équivalent à celui des aminoacyl-ARNt, c'est-à-dire un niveau 10 fois supérieur à celui des ribosomes. (Il s'agit d'un exemple où le surplus de synthèse est partagé entre deux gènes, les gènes, *tufA* et *tufB*.) Il existe une autre différence entre les ribosomes et l'ARN polymérase, qui est en effet présente à un niveau plus faible. Ainsi un mécanisme doit permettre d'augmenter la synthèse des facteurs L7/L12 et EF-Tu et de diminuer celle des sous-unités de l'ARN polymérase comparativement au niveau des protéines ribosomales.

Une caractéristique commune à tous les opérons décrits dans la Figure 10.21 est la régulation autonome de certains gènes par l'un des produits de l'opéron. En

général la protéine régulatrice inhibe l'expression d'un groupe contigu de gènes dans l'opéron, incluant toujours son propre gène. (Le profil exact de l'inhibition varie d'un opéron à un autre.)

Dans chacun des cas, l'accumulation de la protéine entraîne une inhibition de sa propre synthèse et de celle des gènes du groupe dont il fait partie. L'effet se produit souvent au niveau de la traduction de l'ARNm polycistronique et, dans certaines conditions, il peut être reproduit *in vitro*. Ainsi un excès des protéines ribosomales libres entraîne une répression de la traduction.

Chacun des régulateurs est une protéine ribosomale qui interagit directement avec l'ARNr. L'effet au niveau de la traduction résulte de sa capacité à interagir également avec son propre ARNm. Le site d'interaction au niveau de l'ARNm chevauche le site d'initiation de la traduction, ou bien s'étend à une région voisine et influence l'accessibilité du site par des changements conformationnels.

L'utilisation des protéines r, se fixant aux ARNr, pour établir une régulation autonome suggère qu'elles servent de mécanisme de liaison entre leur synthèse et celle des ARNr. Un modèle général est représenté dans la Figure 10.22. Il suppose que les liaisons des protéines r,

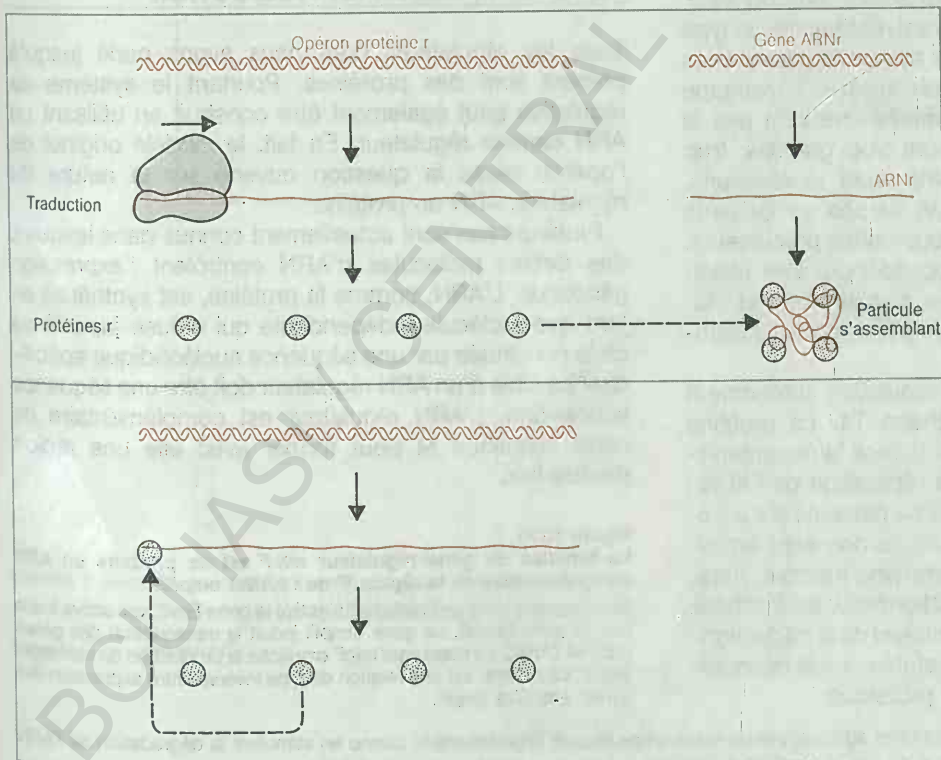


Figure 10.22
La traduction des opérons des protéines r est contrôlée de façon autonome et répond au niveau des ARNr.

En haut : Lorsque les ARNr sont produits, les protéines ribosomales s'associent à eux, de manière à ce qu'il n'existe pas de pool libre de protéines r, et la traduction des ARNm correspondant continue.

En bas : Lorsque tous les ARNr ont été assemblés dans les ribosomes et qu'aucun n'est libre, les protéines r commencent à s'accumuler. L'une d'elles se lie à l'ARNm correspondant et empêche sa traduction.

responsables d'une régulation autonome sont beaucoup plus fortes avec les ARNr qu'avec les ARNm. Ainsi, tant qu'il y aura des ARNr disponibles, les molécules de protéines r s'associeront à eux pour initier l'assemblage des particules ribosomales. Il n'y aura pas de protéine r libre capable de se lier au messager, si bien que la traduction continuera. Toutefois, dès que la synthèse des ARNr ralentit ou s'arrête, des protéines r libres commencent à s'accumuler. Elles deviennent donc capables d'interagir avec leurs ARNm et répriment toute traduction. Ce circuit assure à chaque opéron de protéines r une réponse coordonnée en fonction du niveau des ARNr.

Deux objectifs ont été atteints par ce mode de régulation. Premièrement, le niveau des protéines r est en rapport avec les conditions de croissance de la cellule. En contrôlant le taux des ARNr, la cellule contrôle la production de tous les composants ribosomiaux. Deuxièmement, les autres protéines codées par ces opérons sont synthétisées à leur propre vitesse, indépendamment de la traduction des gènes des protéines r. La sous-unité β de l'ARN polymérase est ainsi soumise à sa propre régulation autonome. Les facteurs EF-Tu et L7/L12 sont traduits avec une plus grande efficacité. Il existe donc, à l'intérieur des circuits de contrôle de ces opérons, des mécanismes permettant des taux de synthèse variables pour des protéines régulées d'une manière coordonnée.

Les régulations autonomes doivent représenter un type de contrôle commun à toutes les protéines incorporées dans des assemblages macromoléculaires. L'avantage est évident ; les particules assemblées ne sont pas le régulateur approprié, car elles sont trop grandes, trop nombreuses et trop limitées dans leurs localisations. Cependant, le besoin de synthèse de ses composants se reflète dans le pool libre des sous-unités précurseurs. Si la voie d'assemblage est bloquée pour une raison quelconque, les sous-unités libres s'accumulent et bloquent la synthèse de tous les composants de l'assemblage qui est devenue inutile.

En fait, le premier exemple de régulation autonome a été fourni par le gène 32 du phage T4. La protéine correspondante joue un rôle central dans la recombinaison génétique, la réparation et la réplication de l'ADN ; sa fonction est la conséquence de sa capacité à s'associer à l'ADN simple-brin. Des mutations non-sens entraînent une forte production de la protéine inactive. Ainsi, lorsque la protéine n'est pas fonctionnelle, sa synthèse est accrue. Cet effet se produit au niveau de la traduction : l'ARNm du gène 32 devient plus stable, il est découplé du comportement de son produit protéique.

Lorsque de l'ADN simple-brin est présent dans la cellule infectée par le phage, il séquestre la protéine p32. En revanche, en absence d'ADN simple-brin, ou dans certaines conditions où la concentration de p32 est excédentaire, la protéine empêche la traduction de son propre ARNm. Nous pouvons supposer que cet effet est la conséquence de la fixation de la protéine sur l'ARNm en empêchant ainsi l'initiation de la traduction. Ceci se produit probablement tout au long d'une région simple-brin proche du site de fixation au ribosome.

Les cellules eucaryotes présentent un autre système de régulation autonome. Le système filamentueux majeur de toutes les cellules eucaryotes est constitué par les microtubules, dont le monomère est la tubuline*. La production de l'ARNm de la tubuline est contrôlée par le pool libre de tubuline. Lorsque ce pool atteint une certaine concentration, la production d'ARNm supplémentaire est bloquée*. Encore une fois, le principe est le même. La protéine constituant l'assemblage moléculaire n'exerce aucune régulation, mais le niveau du précurseur libre détermine si d'autres monomères peuvent être synthétisés.

DES PETITES MOLÉCULES D'ARN PEUVENT RÉGULER LA TRADUCTION

Tous les *régulateurs* dont nous avons parlé jusqu'à présent sont des protéines. Pourtant le système de régulation peut également être construit en utilisant un ARN comme régulateur. En fait, le modèle original de l'opéron laisse la question ouverte sur la nature du régulateur, ARN ou protéine.

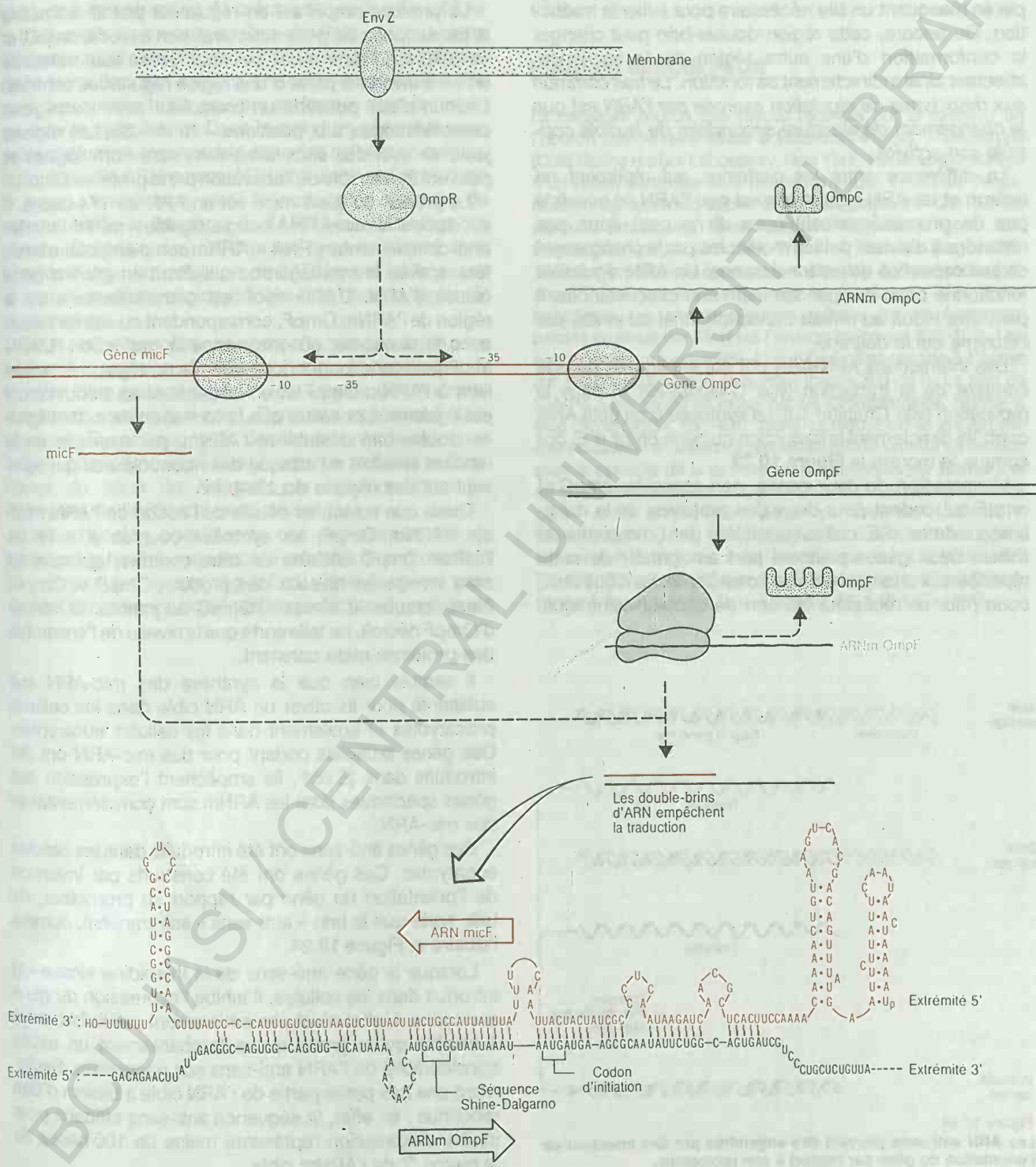
Plusieurs cas sont actuellement connus dans lesquels des petites molécules d'ARN contrôlent l'expression génétique. L'ARN, comme la protéine, est synthétisé en tant que molécule indépendante qui diffuse jusqu'à sa cible constituée par une séquence nucléotidique spécifique. La cible d'un ARN régulateur doit être une séquence simple-brin. L'ARN régulateur est complémentaire de cette séquence et peut former avec elle une région double-brin.

Figure 10.23

La fonction du gène régulateur *micF* est de produire un ARN complémentaire de la région 5' de l'ARNm *ompF*.

Une augmentation de l'osmolarité active le gène *EnvZ*, qui active à son tour le gène *OmpR*. Le gène *OmpR* induit la transcription des gènes *micF* et *OmpC*. Le messager *micF* empêche la production du messager *ompF*. Le résultat est une relation de type inverse entre expression des gènes *ompC* et *ompF*.

* N. d. T. : Les sous-unités libres de tubuline agissent par un mécanisme encore imparfaitement connu en stimulant la dégradation de l'ARN présent dans les polyribosomes. Ce n'est donc pas la production d'ARNm qui serait contrôlée, mais sa stabilité.



La formation d'une région double-brin avec la séquence cible peut empêcher sa fonction directement, par exemple en masquant un site nécessaire pour initier la traduction. Ou encore, cette région double-brin peut changer la conformation d'une autre région de la séquence, affectant ainsi indirectement sa fonction. *Le trait commun aux deux types de régulation exercée par l'ARN est que le changement de structure secondaire de la cible contrôle son activité.*

La différence entre les protéines qui répriment un opéron et les ARN régulateurs est que l'ARN ne possède pas de propriétés allostériques. Il ne peut donc pas répondre à d'autres petites molécules par le changement de ses capacités de reconnaissance. Un ARN régulateur fonctionne donc indépendamment des circonstances. Il peut être induit au niveau transcriptionnel ou inhibé par l'enzyme qui le dégrade.

Des interactions ARN-ARN ont été impliquées dans le contrôle de la traduction (voir Chapitre 29) et dans la réplication (voir Chapitre 13). La synthèse d'un petit ARN contrôle directement la traduction du gène *ompF* d'*E. coli* comme le montre la **Figure 10.23**.

L'expression de deux gènes, non associés, *ompC* et *ompF* qui codent pour deux des protéines de la membrane externe d'*E. coli* est contrôlée par l'osmolarité du milieu. Deux gènes prennent part au contrôle de cette réponse aux changements de l'osmolarité. Le locus *envZ* code pour un récepteur qui sert de capteur osmotique.

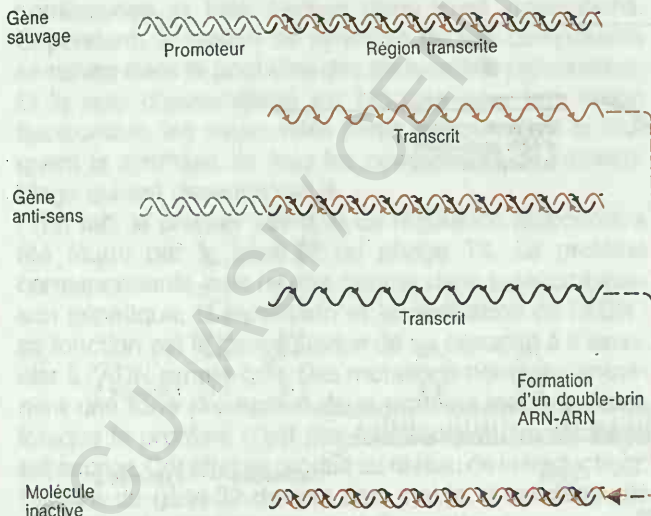


Figure 10.24

Les ARN anti-sens peuvent être engendrés par une inversion de l'orientation du gène par rapport à son promoteur.

Lorsque l'osmolarité augmente la protéine *EnvZ* active une autre protéine, le produit du gène *ompR*.

La protéine *ompR* est un régulateur positif qui active la transcription du gène structural, non associé *ompC*, et du gène régulateur *micF*. Les deux gènes sont transcrits en sens inverse à partir d'une région régulatrice centrale. Chacun d'eux possède un promoteur avec les régions caractéristiques, aux positions - 10 et - 35. Les régions juste en aval des sites d'initiation sont homologues et peuvent être la cible de l'activation par la protéine *OmpR*.

Le produit du gène *micF* est un ARN de 174 bases. Il est appelé le mic-ARN : le sigle signifie « mRNA interfering-complementary RNA », ARNm complémentaire interférant, avec le messager, ce qui décrit en général cette classe d'ARN. L'ARN *micF* est complémentaire de la région de l'ARNm *OmpF*, correspondant au site de liaison avec le ribosome, où commence la traduction. L'ARN *MicF* peut fonctionner donc comme un régulateur, en se liant à l'ARNm *OmpF* et en empêchant sa traduction. Il est également possible que la formation de cette région en double-brin déstabilise l'ARNm, par exemple en le rendant sensible à l'attaque des ribonucléases qui agissent sur des régions double-brin.

Quels que soient les détails de l'action de l'ARN *micF* sur l'ARNm *OmpF*, sa synthèse couplée à celle de l'ARNm *OmpC* entraîne un rétro-contrôle, qui relie en sens inverse les niveaux des produits *OmpF* et *OmpC*. Ainsi, lorsque le niveau d'*OmpC* augmente, le niveau d'*OmpF* décroît, de telle sorte que le niveau de l'ensemble des protéines reste constant.

Il semble bien que la synthèse des mic-ARN soit suffisante pour inactiver un ARN cible dans les cellules procaryotes et également dans les cellules eucaryotes. Des gènes artificiels codant pour des mic-ARN ont été introduits dans *E. coli* ; ils empêchent l'expression des gènes spécifiques dont les ARNm sont complémentaires des mic-ARN.

Des gènes anti-sens ont été introduits dans les cellules eucaryotes. Ces gènes ont été construits par inversion de l'orientation du gène par rapport au promoteur, de telle sorte que le brin « anti-sens » soit transcrit, comme l'illustre la **Figure 10.24**.

Lorsque le gène anti-sens de la thymidine kinase est introduit dans les cellules, il inhibe l'expression du gène endogène. L'effet n'est pas entièrement prévisible, mais il semble qu'un grand excès (probablement un excès considérable) de l'ARN anti-sens soit nécessaire. Seulement une très petite partie de l'ARN cible a besoin d'être reconnue ; en effet, la séquence anti-sens efficace pour inhiber l'expression représente moins de 100 bases de la région 5' de l'ARNm cible.

A quel niveau l'ARN anti-sens inhibe-t-il l'expression ? En principe, il pourrait empêcher soit la transcription du gène, soit la maturation du précurseur, soit enfin inhiber la traduction du messager. Des résultats obtenus dans différents systèmes ont montré que l'inhibition dépendait de la formation des molécules double-brins ARN-ARN, mais ces double-brins peuvent se former dans le noyau ou dans le cytoplasme. Dans le cas d'un gène anti-sens présent d'une manière stable dans des cellules en culture, le double-brin ARNm sens-ARN anti-sens se forme dans le noyau, empêchant ainsi la maturation normale du transcrit et/ou le transport de l'ARN sens dans le cytoplasme. Dans le cas où l'ARNm anti-sens est injecté dans le cytoplasme, la traduction de l'ARNm est inhibée, par la formation d'une région double-brin dans la région 5' du messager.

Cette technique offre une approche très puissante pour inhiber l'expression de gènes ; ainsi, par exemple, la fonction d'un gène régulateur peut être étudiée en introduisant sa version anti-sens. Dans une expérience réalisée avec *D. melanogaster*, des larves ont été obtenues avec le phénotype équivalent à celui de mutants authentiques du gène de *Krüppel*, par injection dans les embryons de l'ARN anti-sens de ce gène.

Une extension possible de ces expériences est de placer un gène anti-sens sous le contrôle d'un promoteur lui-même sujet à une régulation. Ainsi le gène cible devient directement régulé par la production du messager

anti-sens. Cette technique pourrait nous permettre d'étudier l'importance de la cinétique d'expression du gène cible.

POUR EN SAVOIR PLUS

La meilleure source des revues récentes sur la fonction de l'opéron lactose reste **Miller & Reznikoff's** (Eds). *The operons* (Cold Spring Harbor Laboratory, New York, 1978). Les chapitres d'un intérêt particulier sont **Beckwith** (pages 11-30) sur l'organisation du système, **Zabin & Fowler** (pages 89-122) sur les produits protéiques, **Miller** (pages 31-88) sur le gène *lacI*, **Beyreuther** (pages 123-154), **Weber & Geisler** (pages 155-176) sur la protéine répresseur et **Barkley & Bourgeois** (pages 177-220) sur les répresseurs se fixant sur l'ADN et les inducteurs. Les circuits d'autres opérons ont été également passés en revue par **Platt** sur l'opéron *trp* (pages 263-302), **Lee** sur l'opéron *ara* (pages 389-410) et **De Crombrughe & Pastan** sur l'AMPc et l'opéron *galactose* (pages 303-324). Le rôle de l'AMPc a été passé en revue par **De Crombrughe, Busby & Buc** (*Science* 224, 831-838, 1984). La régulation de la synthèse des ribosomes et les mécanismes de rétro-contrôle en relation avec le contrôle de la croissance ont été revus par **Nomura et al.** (*Ann. Rev. Biochem.* 53, 75-117, 1984). Le contrôle « stringent » a été dernièrement passé en revue par **Lamond & Travers** (*Cell* 41, 6-8, 1985). L'utilisation d'ARN anti-sens dans les cellules eucaryotes a été introduit par **Izant & Weintraub** (*Cell* 36 1007-1015, 1984) et a fait l'objet d'une revue par **Green, Pines & Inouye** (*Ann. Rev. Biochem.* 55, 569-597, 1986).

CHAPITRE 11

CONTRÔLE AU NIVEAU DE LA TERMINAISON : L'ATTÉNUATION ET L'ANTITERMINAISON

Une fois que l'ARN polymérase a commencé la transcription, l'enzyme progresse le long de la matrice, et synthétise l'ARN jusqu'à ce qu'elle rencontre une **séquence de terminaison**. A ce signal l'enzyme cesse d'ajouter des nucléotides à la chaîne d'ARN naissante, libère le produit fini, et se dissocie de la matrice d'ADN (on ne sait pas dans quel ordre les deux derniers événements se produisent). La terminaison nécessite que toutes les liaisons hydrogènes liant ensemble l'hybride ARN-ADN soient rompues, après quoi le double-brin d'ADN se reforme.

Il est difficile de déterminer le point de terminaison d'une molécule d'ARN qui a été synthétisée *in vivo* dans une cellule vivante. Il est toujours possible que l'extrémité 3' de la molécule ait été produite par le *clivage* d'un transcrit primaire et que, par conséquent, elle ne représente pas le site réel où l'ARN polymérase s'est arrêtée. Malheureusement, l'extrémité 3' est la même qu'elle soit produite par terminaison ou par clivage. Il n'y a pas de marqueur comparable au triphosphate qui permet d'identifier l'extrémité 5' originale.

La meilleure identification des sites de terminaison est fournie par des systèmes dans lesquels l'ARN polymérase termine la transcription *in vitro*. Comme la capacité de l'enzyme à terminer est fortement influencée par des paramètres tels que la force ionique, sa terminaison à un point particulier *in vitro* ne prouve pas que ce soit le site naturel de terminaison. Mais à présent plusieurs cas sont connus dans lesquels la même extrémité 3' a pu être identifiée sur des espèces particulières d'ARN (procaryotique) indépendamment du fait qu'elles aient été synthétisées *in vivo* ou *in vitro*. Ceci suggère que l'on ait trouvé la séquence naturelle de terminaison.

Les terminateurs dans les bactéries et leurs phages ont été caractérisés avec précision. Ils varient largement à la fois par leurs efficacités de terminaison et leur besoin en protéines auxiliaires, tout du moins *in vitro*. Tous les terminateurs des procaryotes (et ceux de certains eucaryotes) nécessitent la formation d'une « épingle à cheveux » dans la structure secondaire de l'ARN qui est en train d'être transcrit. Dans les bactéries, la capacité à

former ce type de structure secondaire fournit un point précis de régulation de la terminaison.

Au niveau de certains terminateurs, la terminaison peut être évitée par l'intermédiaire de facteurs auxiliaires spécifiques qui interagissent avec l'ARN polymérase. L'**antiterminaison** conduit l'enzyme à passer la séquence de terminaison et à continuer la transcription, un phénomène appelé « **lecture au-delà** » (le même terme est utilisé pour décrire la suppression des codons de terminaison dans le cas des ribosomes).

L'antiterminaison fournit un mécanisme de contrôle élaboré utilisé à la fois par les opérons bactériens et les systèmes régulateurs des phages. Dans les opérons contrôlés par l'**atténuation**, il fournit un lien entre la traduction et la transcription, par lequel la *terminaison de la transcription est inhibée lorsque le ribosome n'est plus capable de progresser le long du segment « leader » de l'ARN_m*. Durant l'infection phagique, différentes protéines auxiliaires (**les facteurs d'antiterminaison**) permettent à l'ARN polymérase de passer des séquences spécifiques de terminaisons. De telles interactions contrôlent l'aptitude de l'enzyme à passer un terminateur et à lire les gènes situés au-delà.

Nous ne devons pas considérer l'événement de terminaison simplement comme un mécanisme de formation de l'extrémité 3' de la molécule d'ARN, mais aussi comme une occasion de contrôler l'expression génique. En effet, les étapes où l'ARN polymérase s'associe avec l'ADN (initiation) ou s'en dissocie (terminaison) peuvent être toutes deux soumises à un contrôle spécifique. Il y a des analogies intéressantes entre les systèmes employés dans l'initiation et dans la terminaison. Toutes deux nécessitent la dissociation des liaisons hydrogènes (d'abord fusion de l'ADN lors de l'initiation, dissociation de l'ARN-ADN lors de la terminaison), toutes deux ont besoin de protéines additionnelles pour interagir avec l'enzyme cœur. En fait, les deux processus peuvent être accomplis par des formes différentes de l'enzyme.

DEUX MODES DE TERMINAISON DANS *E. COLI* METTENT EN JEU DES PALINDROMES

Les séquences au niveau des terminateurs procaryotiques ne montrent pas d'homologie au-delà du point au niveau duquel la dernière base est ajoutée à l'ARN. Ainsi la responsabilité de la terminaison peut être imputée aux *séquences déjà transcrites* par l'ARN polymérase. La terminaison repose donc sur l'examen de la région que la polymérase est en train de transcrire.

Tous les terminateurs procaryotiques contiennent des séquences palindromiques juste avant le point de termi-

naison. Dans chaque cas, le palindrome est constitué de répétitions inversées séparées par une courte distance. L'appariement de ces répétitions inversées peut engendrer dans l'ARN des épingles à cheveux constituées de structures en boucle et en tige, comme cela est illustré dans la **Figure 11.1**.

Les mutations ponctuelles qui empêchent la terminaison apparaissent dans une région d'environ 35 pb avant le site réel de terminaison. Elles se trouvent toutes dans la région en tige de l'épingle à cheveux. La plupart rompent l'appariement des bases, mais pas toutes ; la stabilité de la région en tige de l'épingle à cheveux ne peut donc pas être l'unique facteur impliqué dans la terminaison. (Toutefois, les mutations qui favorisent l'appariement *augmentent* en général la terminaison.)

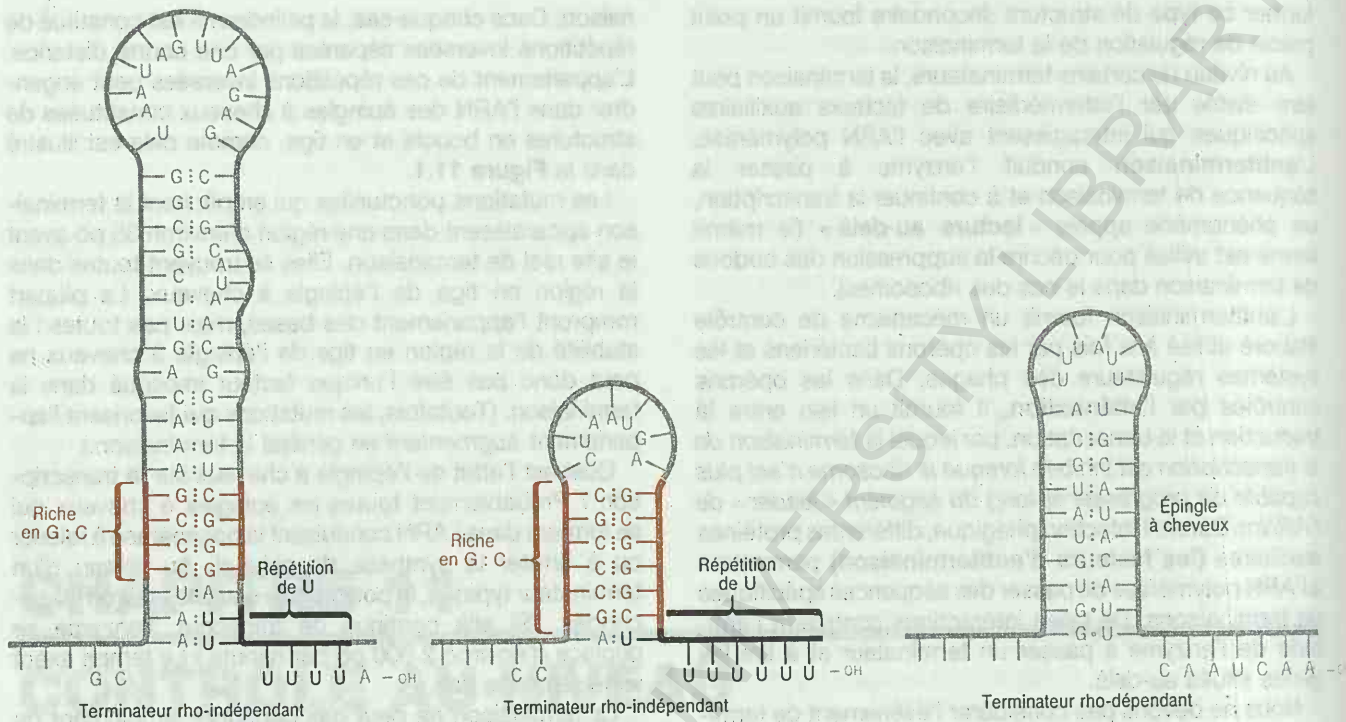
Quel est l'effet de l'épingle à cheveux sur la transcription ? Probablement toutes les épingles à cheveux qui se forment dans l'ARN conduisent la polymérase à ralentir ou à arrêter la synthèse de celui-ci. Au niveau d'un terminateur typique, la polymérase s'arrête environ 60 secondes. (Si elle continue de transcrire, l'enzyme se déplace d'environ 2 000 pb par minute.) Le temps exact varie selon les cas.

La terminaison ne peut pas dépendre simplement de la rencontre d'une épingle à cheveux ; il y a trop de structures secondaires dans l'ARN pour que ceci soit vraisemblable. Les événements qui suivent l'arrêt à un palindrome doivent dépendre des séquences voisines. Un arrêt augmente la probabilité d'apparition de la terminaison. Mais si aucune séquence de terminaison n'est présente, l'enzyme reprend en général la transcription.

Les séquences terminatrices d'*E. coli* ont été classées suivant que l'ARN polymérase exige ou non des facteurs additionnels pour terminer la transcription *in vitro* :

- L'enzyme cœur permet de terminer *in vitro* la transcription à certains sites terminateurs en l'absence de tout autre facteur. Ces sites sont appelés **rho-indépendants** ou **terminateurs simples**. Il est tout à fait possible que d'autres facteurs soient nécessaires *in vivo*.
- Les terminateurs **rho-dépendants** sont définis par leur besoin en **facteur rho** *in vitro* ; les mutations montrent que le facteur est aussi impliqué dans la terminaison *in vivo*.

Les sites de terminaisons simples ont deux caractéristiques structurales : l'épingle à cheveux commune à tous les sites de terminaisons, et une queue d'environ 6 résidus U située à l'extrémité de l'unité de transcription. Ces deux caractéristiques sont nécessaires à la terminaison.

**Figure 11.1**

Les séquences des terminateurs contiennent des régions palindromiques qui forment des épingles à cheveux de longueur variant de 7 à 20 pb. Les structures en tige et boucle sont suivies par la répétition de résidus U dans les sites rho-indépendants mais pas dans les sites rho-dépendants.

Les chaînes de résidus U fournissent probablement un signal qui permet à l'ARN polymérase de se dissocier de la matrice quand elle s'arrête à une épingle à cheveux. L'hybride rU-dA ARN-ADN possède une structure en bases appariées particulièrement faible ; c'est, parmi les hybrides ARN-ADN, celui dont la dissociation nécessite l'énergie la plus faible. Quand l'ARN polymérase s'arrête, l'hybride ARN-ADN peut se dissocier à partir des liaisons faibles rU-dA de la région terminale. Souvent les événements de terminaisons réels se produisent à n'importe quelle position vers ou à la fin de l'extrémité U, comme si l'enzyme « bégayait » pendant la terminaison.

L'importance de la répétition de U est confirmée en effectuant des délétions qui enlèvent certains U : bien que la polymérase s'arrête toujours au niveau de l'épingle à cheveux, elle ne termine plus la transcription. La série de bases U correspond naturellement à une région de l'ADN riche en A-T ; nous voyons donc que les régions riches en A-T peuvent être importantes dans la terminaison simple aussi bien que dans l'initiation.

Lorsque certaines unités de transcription sont utilisées comme matrice dans la synthèse d'ARN *in vitro*, la terminaison correcte n'apparaît pas ; la polymérase s'arrête au site de terminaison, mais ensuite poursuit la synthèse de l'ARN. Il a été trouvé que le facteur rho est une protéine dont l'addition dans les systèmes *in vitro* permet à l'ARN polymérase de terminer à certains sites en engendrant des molécules d'ARN avec des extrémités 3' particulières. Cette fonction est appelée la terminaison **rho-dépendante**.

Le besoin en facteur rho *in vitro* est toutefois variable. Certains terminateurs nécessitent de relativement hautes concentrations de rho alors que d'autres fonctionnent tout aussi bien à plus faibles concentrations. Nous ne savons pas si ce type d'effet représente un artefact du système *in vitro* ou s'il représente une véritable différence dans l'efficacité de la terminaison *in vivo*.

Les terminateurs rho-dépendants ont des épingles à cheveux qui ne sont pas particulièrement riches en G-C ; aucune séquence commune n'a pu être identifiée à

l'intérieur ou près de l'épingle à cheveux. Il n'y a pas de répétitions de U ni d'autre région riche en A-T.

LA TERMINAISON CHEZ LES EUCARYOTES MET AUSSI EN JEU DES STRUCTURES SECONDAIRES OU DES RÉPÉTITIONS DE U

Il y a peu de choses connues au sujet des signaux de terminaisons ou des procédés utilisés par la plupart des ARN polymérases eucaryotes. La principale difficulté dans l'analyse des transcrits est le manque de certitude concernant les véritables sites de terminaisons. Bien que les molécules d'ARN identifiées *in vivo* aient des extrémités 3' précises, comment peut-on savoir si elles ont été produites par terminaison ou par coupure d'un transcrit primaire plus long ?

Pour les produits de l'ARN polymérase II, le problème est exacerbé par la modification qui se produit au niveau de l'extrémité 3' et qui consiste en l'addition d'une longue séquence d'acide polyadénylique. L'extrémité 3' pour la polydénylation est habituellement engendrée par coupure et non par terminaison (voir Chapitre 24). La région 3' du transcrit primaire d'origine demeure généralement non caractérisée, mais nous avons cependant des informations sur les signaux nécessaires à la formation de l'extrémité à laquelle est ajouté l'acide polyadénylique.

Il est possible que la terminaison par l'ARN polymérase II soit induite seulement de façon assez floue. Dans certaines unités de transcription la terminaison a lieu à plus de 1 000 pb en aval du site correspondant à l'extrémité 3' de l'ARNm mature (qui est engendrée par coupure au niveau d'une séquence spécifique). Au lieu d'utiliser des séquences terminatrices spécifiques, l'enzyme peut cesser la synthèse de l'ARN au niveau de sites multiples situés à l'intérieur d'une région terminatrice relativement grande. La nature des sites distincts de terminaison n'est pas connue.

Un site de terminaison a été identifié dans l'ADN du virus SV40 ; cette séquence contient une épingle à cheveux suivie d'une répétition de bases U semblable à un site *t* bactérien rho-indépendant. Une séquence similaire est trouvée à la fin de la région terminatrice de la globine.

On a montré l'importance de la structure secondaire dans la production des extrémités 3' en analysant la capacité de matrices variées à donner naissance à des ARNm d'histones (non polyadénylés) quand les matrices sont injectées dans des oocytes de xénope. L'ARN est achevé au niveau d'une structure en tige et boucle, et les mutations qui empêchent la formation de cette tige double brin inhibent la terminaison de l'ARN. Les muta-

tions secondaires qui restaurent la structure bicaténaire (et pas nécessairement la séquence d'origine) se comportent comme des révertants. Ceci suggère que la formation des structures secondaires est plus importante que la séquence exacte.

En principe, chacun des brins ou les deux brins de l'ADN pourraient être impliqués dans la formation de structures secondaires. Ils peuvent être différenciés en utilisant des matrices constituées de double-brins hétérologues dans lesquels les deux brins de l'ADN ne sont pas identiques. Il s'est avéré qu'il est important de pouvoir former une structure bicaténaire pour le brin codant, mais pas pour le brin qui est utilisé comme matrice. Ceci suggère que la structure secondaire produit son effet en se formant dans l'ARN au fur et à mesure qu'il est transcrit.

Pour l'ARN polymérase I, l'unique produit est un grand précurseur qui contient la majorité des séquences des ARNr. Le précurseur est soumis à une maturation complexe. La terminaison se fait à un site spécifique situé à plus de 1 000 pb en aval de l'extrémité 3', qui est probablement engendrée par clivage.

Avec l'ARN polymérase III, la transcription *in vitro* produit des molécules ayant les mêmes extrémités 3' et 5' que celles synthétisées *in vivo*. Ceci suggère que l'initiation et la terminaison ont été correctes. Le système peut être manipulé en introduisant des changements de séquences dans la matrice aux alentours de la région de terminaison.

Quand les gènes 5S de *X. laevis* sont transcrits *in vitro* par une enzyme homologue, la terminaison se fait au niveau de répétitions de 4 bases U. Habituellement, la terminaison survient à la 2^e base, mais il existe une hétérogénéité, certaines molécules se terminant à la 3^e voire même à la 4^e base U. On trouve la même hétérogénéité dans les molécules synthétisées *in vivo* ; il semble donc que ce soit un trait caractéristique de la réaction de terminaison.

Tout comme dans les terminateurs procaryotiques, la répétition de U est enchâssée dans une région riche en G-C. Bien que des séquences à symétrie « dyad » soient présentes, elles ne sont pas nécessaires à la terminaison, puisque des mutations qui abolissent la symétrie n'empêchent pas l'achèvement normal de la synthèse d'ARN. Aucune séquence au-delà de la répétition de U n'est nécessaire, puisque tous les nucléotides à sa droite peuvent être remplacés par d'autres séquences sans avoir d'effet sur la terminaison.

La répétition de U à elle seule n'est pas suffisante pour la terminaison, puisqu'il existe des régions de 4 résidus U successifs à l'intérieur des unités de transcription lues par l'ARN polymérase III. (Cependant il n'y a pas de

final. Ainsi quand la cellule a suffisamment de tryptophane, elle est capable d'arrêter la synthèse de futures molécules d'acides aminés par inhibition au début de la voie métabolique.

Une carence du répresseur produit une brusque augmentation d'environ 70 fois de la fréquence des événements d'initiation au niveau du promoteur *trp*. Même dans les conditions de répression, les gènes de structure continuent à être exprimés à un faible niveau **basal** dit **niveau réprimé**. Ceci veut dire que l'efficacité de la répression au niveau de l'opérateur est beaucoup plus faible que dans l'opéron *lac* (où le niveau basal est seulement 1/1 000 du niveau induit).

En plus du complexe promoteur-opérateur, un autre site est impliqué dans la régulation de l'opéron *trp*. Son existence a été mise en évidence pour la première fois en observant que la délétion d'une séquence située entre l'opérateur et la région codant pour *trpE* pouvait augmenter l'expression des gènes de structure. Cet effet est indépendant de la répression : à la fois le niveau basal et le niveau déréprimé sont augmentés. Par conséquent, ce site influence des événements qui interviennent *après* que l'ARN polymérase ait quitté le promoteur (quelles que soient les conditions ayant prévalu à l'initiation).

Le site de régulation est appelé **atténuateur**. Il est situé à l'intérieur de la séquence leader transcrite de 162 nucléotides qui précède le codon initiateur du gène *trpE*. *L'atténuateur est un obstacle à la transcription*. C'est un site de terminaison rho-indépendant : un petit palindrome riche en G-C est suivi de huit résidus U successifs. L'ARN polymérase s'y arrête soit *in vivo*, soit *in vitro* et fournit un transcrit de 140 bases.

La terminaison à ce site dépend du niveau de tryptophane, comme cela est illustré dans la **Figure 11.3**. En présence d'une quantité suffisante de tryptophane, la terminaison est effective. Mais en l'absence de tryptophane, l'ARN polymérase peut continuer dans les gènes de structure. Ce type de régulation des événements de terminaison est appelé **atténuation**.

L'atténuation contrôle l'opéron dans le même sens que la répression. Quand le tryptophane est présent, l'opéron est réprimé ; et la plupart des ARN polymérases qui échappent au promoteur s'arrêtent à l'atténuateur. Quand le tryptophane est retiré, l'ARN polymérase a accès libre au promoteur et n'est plus contrainte à terminer prématurément.

L'atténuation permet une variation d'un facteur 8 à 10 de la transcription. Dans les conditions de terminaison effective, environ 10 % des ARN polymérases qui rencontrent le terminateur peuvent continuer. Dans les condi-

tions d'atténuation, la plupart des ARN polymérases peuvent continuer la transcription. Avec l'augmentation de 70 fois de l'initiation de la transcription due à la levée de la répression, ceci permet une variation de la régulation de l'opéron d'un facteur d'environ 600.

Comment la terminaison de la transcription à l'atténuateur peut-elle dépendre du niveau du tryptophane ? La séquence de la région « leader » suggère un mécanisme.

La **Figure 11.4** montre qu'elle contient un site de fixation des ribosomes dans lequel le codon AUG est suivi par une région codante de 13 codons. Cette séquence est-elle traduite en un **peptide leader** ? Bien qu'aucun produit n'ait été détecté *in vivo*, c'est probablement parce qu'il est instable. Nous savons que le site de fixation du ribosome est fonctionnel (car, lorsqu'il est fusionné à un gène de structure, il donne une traduction efficace).

Quelle est la fonction du peptide leader ? Il contient deux résidus tryptophane qui se suivent. Le tryptophane est un acide aminé rare dans les protéines d'*E. coli* ; il est donc peu probable que ce soit une pure coïncidence. Quand la cellule a épuisé son tryptophane, les ribosomes amorcent la traduction du peptide « leader », mais s'arrêtent quand ils atteignent les codons Trp. La séquence de l'ARNm suggère que **l'arrêt du ribosome** peut influencer la terminaison au niveau de l'atténuateur.

La séquence leader peut être écrite sous la forme d'appariements alternatifs. La capacité du ribosome à poursuivre à travers la région leader peut contrôler la transition entre ces structures. La structure détermine si l'ARNm peut fournir les éléments nécessaires à la terminaison.

La **Figure 11.5** présente ces structures. Tout d'abord la région 1 s'apparie avec la région 2 et la région 3 avec la région 4. L'appariement des régions 3 et 4 engendre l'épingle à cheveux qui précède la séquence U_8 : c'est le signal essentiel de terminaison. L'ARN prendrait probablement cette structure en présence d'aucune intervention extérieure.

Une structure différente est formée si la région 1 est empêchée de s'apparier avec la région 2. Dans ce cas la région 2 est libre de s'apparier avec la région 3. Alors la région 4 n'a plus de partenaire disponible pour s'apparier. De cette façon elle est contrainte à rester en simple brin. Et donc l'épingle à cheveux, signal de terminaison, ne peut plus être formée.

La **Figure 11.6** montre que la position du ribosome peut déterminer quelle structure est formée, de telle sorte que la terminaison soit atténuée uniquement en absence de tryptophane.

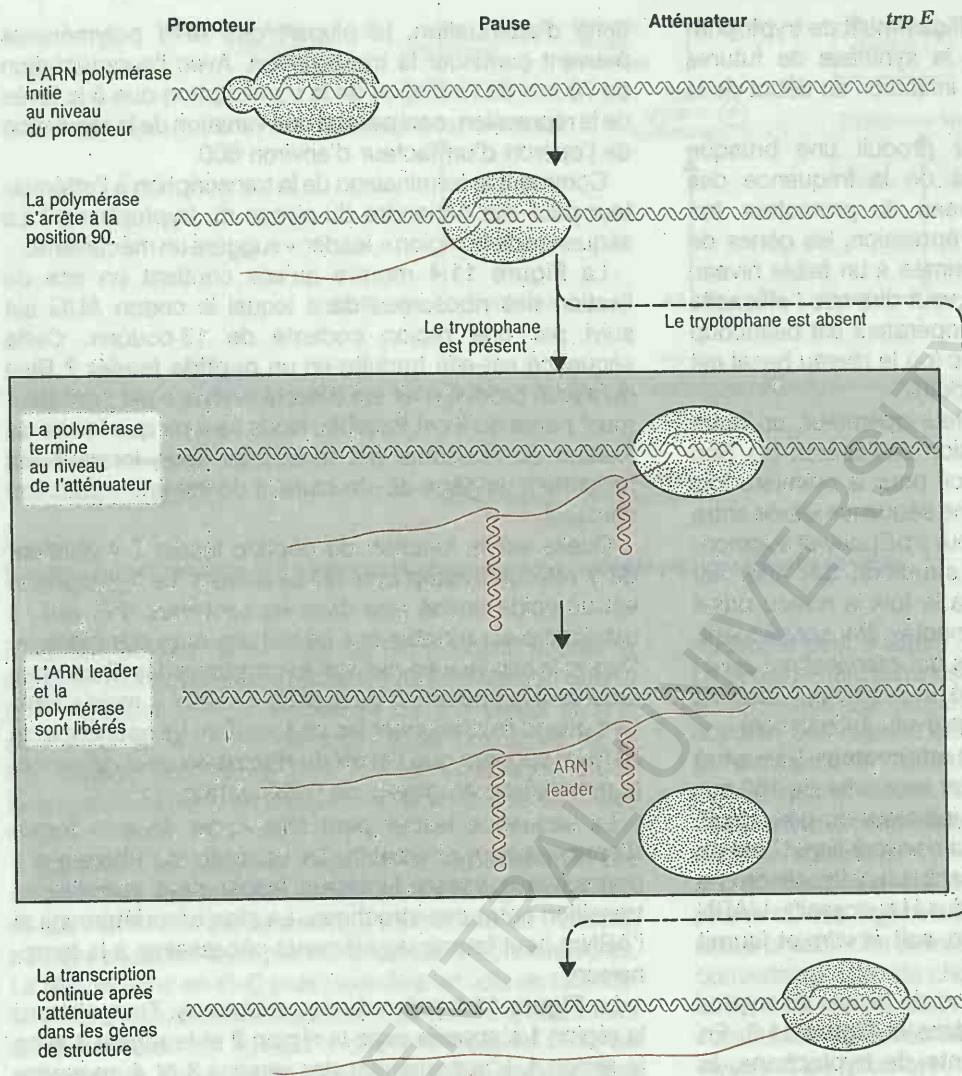


Figure 11.3

Un atténuateur contrôle la progression de l'ARN polymérase dans les gènes *trp*.

Après l'initiation au niveau du promoteur (en condition d'expression basale ou de dérégulation) l'ARN polymérase progresse jusqu'à la position 90, où elle s'arrête. Quand elle atteint l'atténuateur en présence de tryptophane, il y a 90 % de chance que la terminaison se produise et que l'ARN leader de 140 bases soit libéré. En l'absence de tryptophane, la polymérase continue dans les gènes de structure (*trpE* commence à la position + 163).

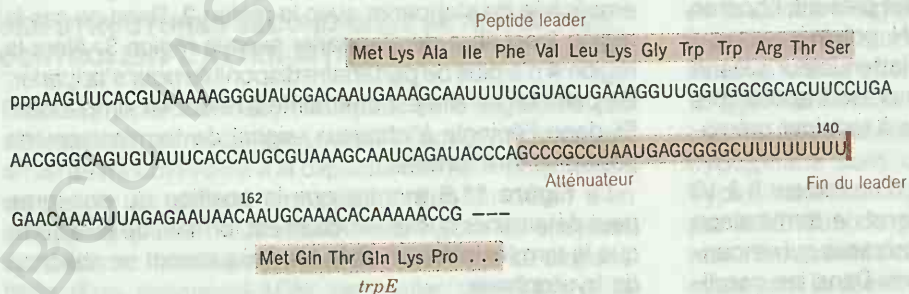


Figure 11.4

La séquence leader des gènes *trp* code pour un peptide court.

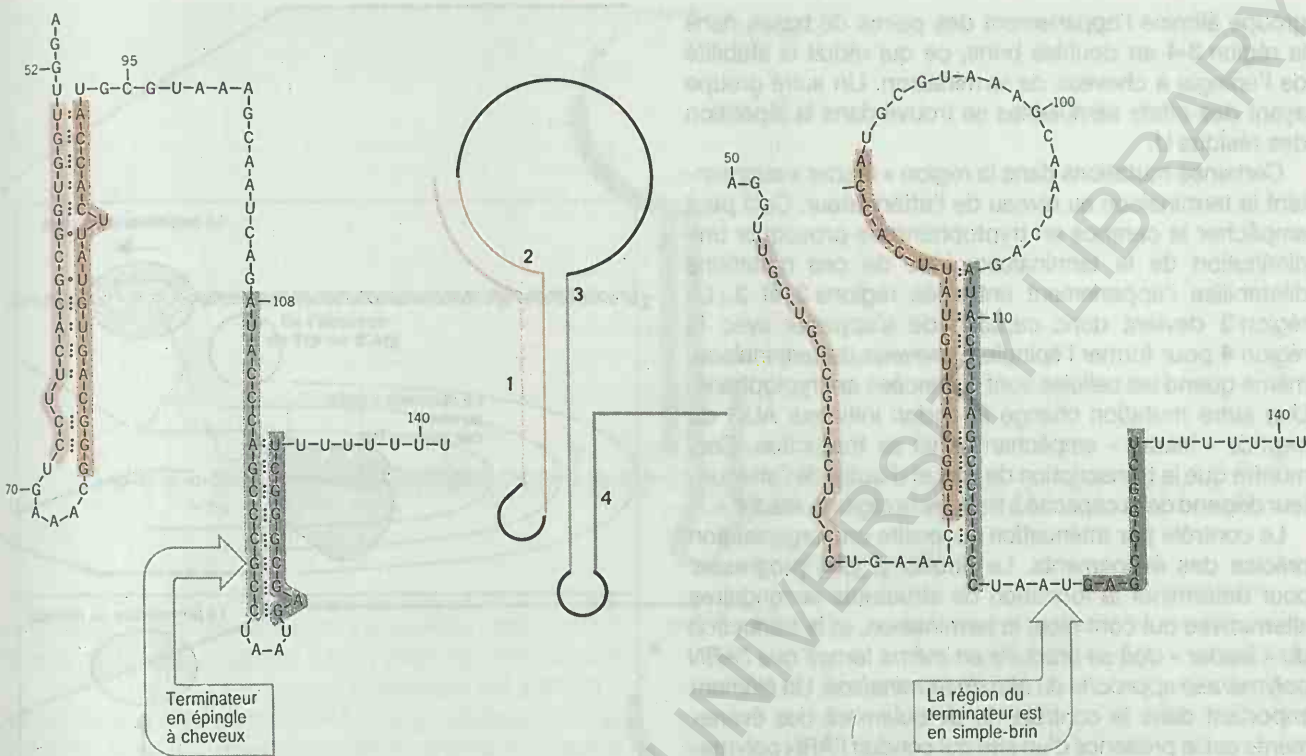


Figure 11.5
La région leader du *trp* peut exister dans des conformations de bases appariées alternatives.

Dans le centre, se trouve le diagramme représentant les quatre régions qui peuvent s'apparier. La région 1 est complémentaire de la région 2, qui est elle-même complémentaire de la région 3, elle-même complémentaire de la région 4. À gauche, se trouve la conformation obtenue, quand la région 1 s'apparie avec la région 2 et que la région 3 s'apparie avec la région 4. Sur la droite, se trouve la conformation obtenue quand la région 2 s'apparie avec la région 3, laissant les régions 1 et 4 non appariées.

Quand le tryptophane est présent, les ribosomes sont capables de synthétiser le peptide « leader ». Ils continueront le long de la section leader de l'ARNm jusqu'au codon UAG, qui se trouve entre les régions 1 et 2. Comme le montre la figure, en progressant vers ce point, le ribosome masque la région 2 et l'empêche de s'apparier. Le résultat est que la région 3 est libre de s'apparier avec la région 4, engendrant le terminateur en épingle à cheveux. Dans ces conditions, donc, l'ARN polymérase s'arrête à l'atténuateur.

Quand il n'y a pas de tryptophane, les ribosomes s'arrêtent aux codons *trp*, qui font partie de la région 1. Ainsi la région 1 est masquée par le ribosome et ne peut plus s'apparier avec la région 2. Si cet événement survient alors que l'ARNm est en train d'être synthétisé les

régions 2 et 3 s'apparieront avant même que la région 4 ait été transcrite. Ceci contraint la région 4 à rester sous forme de simple brin. En l'absence d'épingle à cheveux de terminaison, l'ARN polymérase passe l'atténuateur et continue la transcription.

La carence en d'autres acides aminés n'a pas cet effet, car les sites où s'arrête le ribosome laissent les régions 1 et 2 susceptibles de s'apparier ; ainsi les régions 3 et 4 peuvent s'apparier pour former l'épingle à cheveux de terminaison.

Les mutations qui modifient l'atténuation agissent par leurs effets sur la terminaison. Les plus fréquentes produisent un défaut de terminaison, augmentant ainsi l'expression des gènes de structure. Ces mutations sont similaires à celles trouvées dans d'autres terminateurs. Un premier

groupe élimine l'appariement des paires de bases dans la région 3-4 en doubles brins, ce qui réduit la stabilité de l'épingle à cheveux de terminaison. Un autre groupe ayant des effets semblables se trouve dans la répétition des résidus U.

Certaines mutations dans la région « leader » augmentent la terminaison au niveau de l'atténuateur. Ceci peut empêcher la carence en tryptophane de provoquer une diminution de la terminaison. Une de ces mutations déstabilise l'appariement entre les régions 2 et 3. La région 3 devient donc capable de s'apparier avec la région 4 pour former l'épingle à cheveux de terminaison, même quand les cellules sont carencées en tryptophane. Une autre mutation change le codon initiateur AUG du peptide « leader » empêchant ainsi sa traduction. Ceci montre que la transcription de part et d'autre de l'atténuateur dépend de la capacité à traduire la région « leader ».

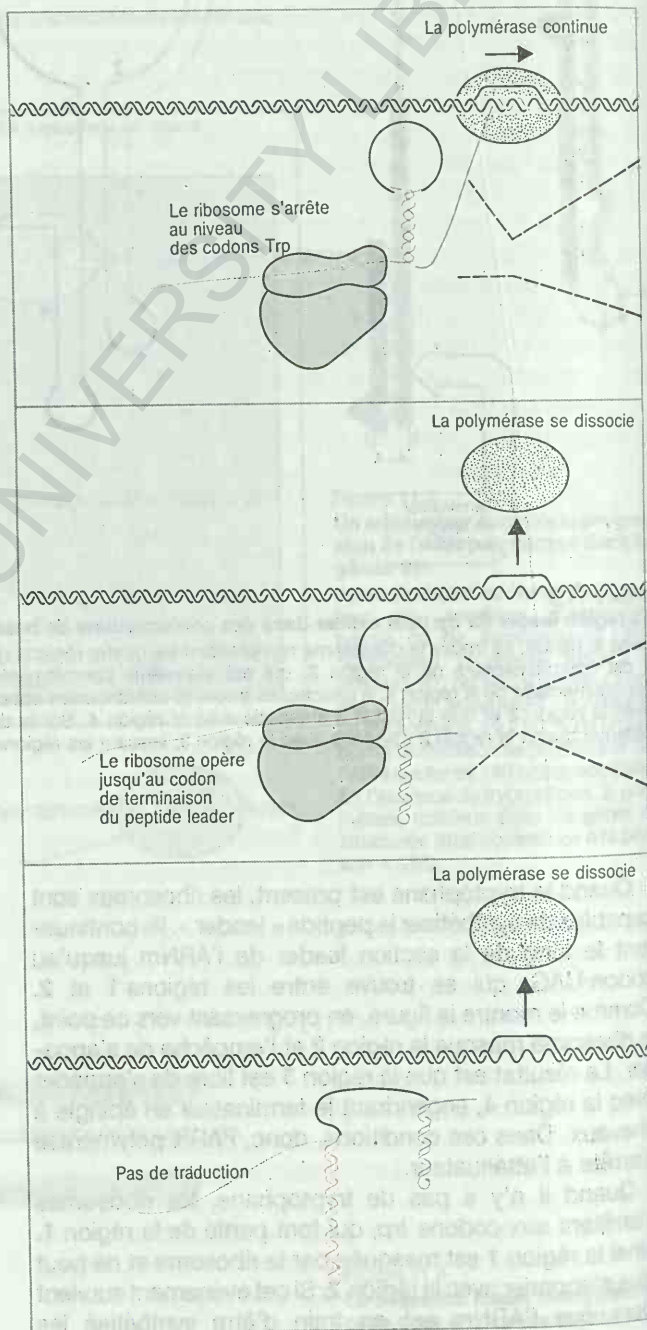
Le contrôle par atténuation nécessite une organisation précise des événements. Le ribosome doit progresser pour déterminer la formation de structures secondaires alternatives qui contrôlent la terminaison, et la *traduction du « leader » doit se produire en même temps que l'ARN polymérase approche du site de terminaison*. Un élément important dans le contrôle du déroulement des événements est la présence d'un site qui conduit l'ARN polymérase à faire une pause à la base 90 du peptide « leader ». L'ARN polymérase reste là jusqu'à ce que le ribosome traduise le peptide leader, puis la polymérase redémarre et se dirige vers le site d'atténuation. Au moment où la polymérase arrive, les structures secondaires de la région d'atténuation ont été fixées.

L'atténuation répond directement au besoin de la cellule en tryptophane nécessaire à la synthèse des protéines, en fournissant un mécanisme de détection des anomalies d'approvisionnement en ARNt trp.

LA PORTÉE GÉNÉRALE DE L'ATTÉNUATION

L'utilisation de l'atténuation comme mécanisme de contrôle des opérons bactériens est-elle très répandue ? Elle est utilisée par au moins six opérons qui codent pour des enzymes impliquées dans la biosynthèse des acides aminés. Ainsi un rétrocontrôle de la production des enzymes de synthèse des acides aminés par le niveau de l'acide aminé disponible pour la synthèse doit être fréquent.

Les séquences des peptides leaders des opérons contrôlés par atténuation sont résumées dans la **Figure 11.7**. Chaque arrêt du ribosome au codon représentant



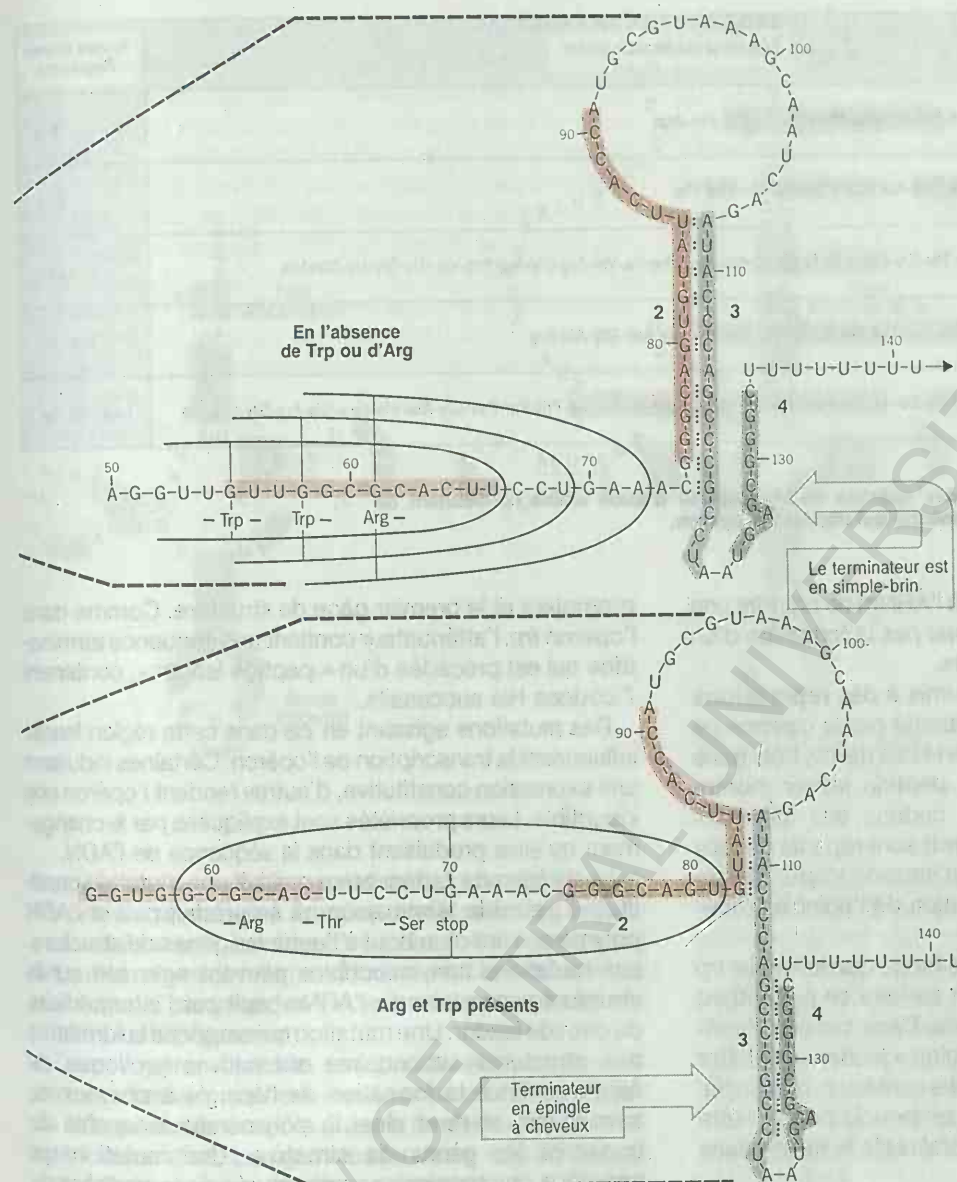


Figure 11.6

Les alternatives de l'ARN polymérase au niveau de l'atténuateur, dépendent de la localisation du ribosome, lequel détermine si les régions 3 et 4 peuvent s'apparier pour former le terminateur en épingle à cheveux.

Opéron	Séquence de peptide leader	Acides aminés Régulateur
his	Met-Thr-Arg-Val-Gln-Phe-Lys-His-His-His-His-His-His-His-Pro-Asp	His
pheA	Met-Lys-His-Ile-Pro-Phe-Phe-Phe-Ala-Phe-Phe-Phe-Thr-Phe-Pro	Phe
leu	Met-Ser-His-Ile-Val-Arg-Phe-Thr-Gly-Leu-Leu-Leu-Leu-Asn-Ala-Phe-Ile-Val-Arg-Gly-Arg-Pro-Val-Gly-Gly-Ile-Gln-His	Leu
thr	Met-Lys-Arg-Ile-Ser-Thr-Thr-Ile-Thr-Thr-Thr-Ile-Thr-Ile-Thr-Thr-Gly-Asn-Gly-Ala-Gly	Thr Ile
ilv	Met-Thr-Ala-Leu-Leu-Arg-Val-Ile-Ser-Leu-Val-Val-Ile-Ser-Val-Val-Val-Ile-Ile-Ile-Pro-Pro-Cys-Gly-Ala-Ala-Leu-Gly-Arg-Gly-Lys-Ala	Leu, Val, Ile

Figure 11.7

Les séquences de peptides leader des opérons de biosynthèse d'acide aminé contiennent de multiples codons d'(des) acide(s) aminé(s) qui régule(nt) l'opéron.

l'acide aminé régulateur permet à l'ARNm de prendre une structure secondaire qui ne permet pas la formation d'un terminateur en épingle à cheveux.

Les opérons *thr* et *ilv* sont soumis à des **répressions multivalentes**. Chacun est dérprimé par la carence de plus d'un acide aminé, deux dans le cas de *thr*, trois dans le cas de *ilv*. La séquence du peptide leader montre comment cela se produit : les codons des différents acides aminés qui régulent l'opéron sont répartis de telle sorte que l'arrêt du ribosome à n'importe lequel d'entre eux permet d'empêcher la formation de l'épingle à cheveux de terminaison.

Plusieurs des leaders sont plus longs que le leader *trp* et les positions où la pause peut se faire en permettant la dérpression sont plus étendues. Dans ces cas, l'arrêt de plus d'un ribosome dans la région « leader » peut être nécessaire pour aboutir à une dérpression maximale. Plusieurs structures secondaires en boucle peuvent être impliquées, mais le principe général reste le même dans chaque cas.

L'opéron histidine contient un groupe de 9 gènes de structure représentant les enzymes qui synthétisent l'histidine à partir du PRPP. La carence en histidine provoque une augmentation de sa transcription d'un facteur 10, qui est en relation inverse avec la quantité d'His-ARNt dans la cellule. L'opéron peut être dérprimé par la mutation de n'importe lequel des différents gènes, ce qui entraîne dans tous les cas une inhibition de la production d'His-ARNt fonctionnel.

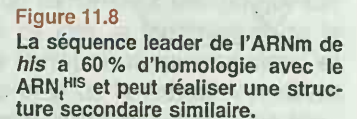
Cette situation est expliquée par la présence d'un atténuateur dans la région « leader » qui s'étend entre le

promoteur et le premier gène de structure. Comme dans l'opéron *trp*, l'atténuateur contient une séquence terminatrice qui est précédée d'un « peptide leader », contenant 7 codons His successifs.

Des mutations agissant en *cis* dans cette région leader influencent la transcription de l'opéron. Certaines induisent une expression constitutive, d'autres rendent l'opéron non inductible. Leurs propriétés sont expliquées par le changement qu'elles produisent dans la séquence de l'ADN.

La délétion de l'atténuateur produit une mutation constitutive ; ainsi la terminaison ne se produit pas et l'ARN polymérase lira d'un bout à l'autre les gènes de structure. Les mutations non inductibles peuvent agir, soit sur la structure secondaire de l'ARNm, soit par l'intermédiaire du peptide leader. Une mutation qui empêche la formation des structures secondaires alternatives provoque de *façon continue* la formation de l'épingle à cheveux de terminaison et rend ainsi la polymérase incapable de transcrire les gènes de structure. Une mutation qui provoque une terminaison prématurée de la synthèse du peptide leader crée effectivement une situation de non-traduction similaire à celle qui est montrée dans la Figure 11.6 dans laquelle l'épingle de terminaison se forme en permanence.

Les similarités entre les mécanismes d'atténuation du tryptophane et de l'histidine sont marquées. Dans chaque cas, c'est la carence en aminoacyl-ARNt qui prévient directement la terminaison de la transcription et ainsi entraîne la transcription des gènes de structure. L'opéron *trp* a aussi une régulation par interaction opérateur-répresseur, alors que dans l'opéron histidine l'atténuation



L'atténuation dépend de la capacité de structures secondaires dans l'ARNm à déterminer si l'ARN polymérase reconnaît un site rho-indépendant, au niveau duquel elle arrêtera la transcription. La structure secondaire est également importante dans la terminaison rho-dépendante, laquelle fournit une cible pour la régulation, durant

l'infection phagique et aussi, peut-être, pour la bactérie non infectée elle-même.

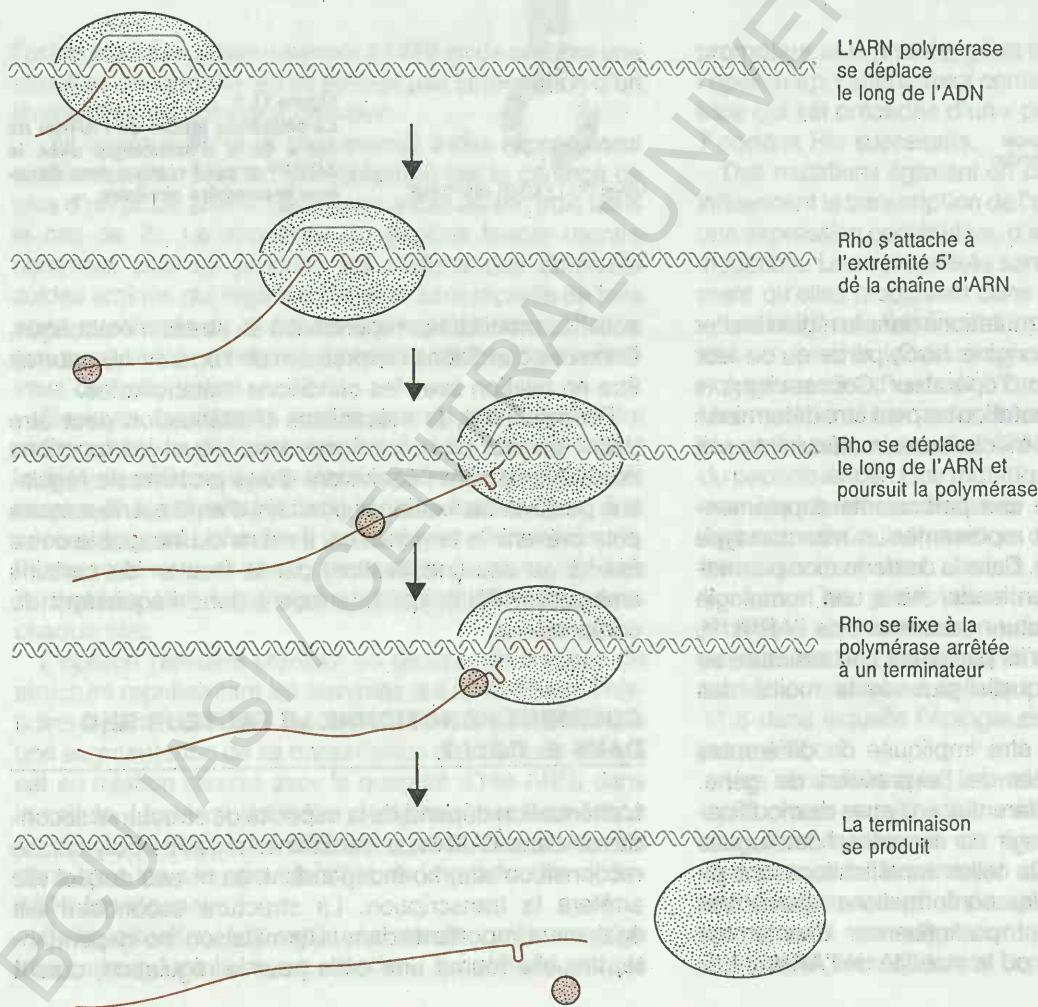
Le facteur rho fonctionne uniquement à l'étape de terminaison. C'est une protéine d'environ 55 000 daltons, probablement active sous forme de multimère. Il manifeste deux activités *in vitro*. Il fonctionne catalytiquement comme un facteur de terminaison ; normalement son maximum d'activité est obtenu quand il est présent à environ 10 % de la concentration de l'ARN polymérase. Il a également une activité NTPase — ce qui veut dire qu'il hydrolyse les nucléotides triphosphates libres — ce qui est nécessaire à la terminaison.

Le facteur rho agit-il en reconnaissant l'ADN, l'ARN ou l'ARN polymérase ? Son activité NTPase est ARN-dépendante. Elle requiert idéalement la présence d'un polyribonucléotide de plus de 50 résidus. Ceci suggère que rho se fixe à l'ARN. Un modèle d'action de rho prévoit qu'il se fixe à l'extrémité 5' d'une chaîne d'ARN naissante, puis qu'il progresse le long de l'ARN, utilisant l'hydrolyse des NTP pour fournir l'énergie nécessaire au déplacement.

La Figure 11.9 suggère que l'arrêt de l'ARN polymérase à ce site de terminaison donne une opportunité à rho de s'y fixer. Puis le facteur interagit avec l'ARN polymérase,

Figure 11.9

Le facteur rho poursuit l'ARN polymérase le long de l'ARN et peut provoquer la terminaison s'il se fixe à l'enzyme arrêtée au niveau d'un terminateur rho-dépendant.



la conduisant à libérer l'ARN ; les deux protéines se dissocient alors de l'acide nucléique.

Ce modèle prévoit que le point critique du site de terminaison est sa capacité à produire une pause de l'ARN polymérase. Les différences entre les différents terminateurs peuvent être expliquées s'ils varient dans l'efficacité avec laquelle ils induisent la pause, parce qu'ils retardent la polymérase pour des durées caractéristiques différentes. Le retard peut dépendre de la stabilité de l'épingle à cheveux, qui est fonction de sa longueur et de sa richesse en G-C. Nous supposons habituellement que rho n'agit pas à n'importe quel site où l'ARN polymérase s'est arrêtée, mais qu'il exige des séquences spécifiques supplémentaires.

L'idée que rho se déplace le long de l'ARN conduit à une hypothèse importante concernant les relations entre la transcription et la traduction. Rho doit être le premier à avoir accès à l'extrémité 5' libre pour être ensuite capable de se déplacer le long de l'ARN. Ces événements sont probablement supprimés si les ribosomes sont en train de traduire l'ARN. De cette façon l'aptitude du facteur rho à atteindre le terminateur *tandis que l'ARN polymérase y est arrêtée* doit dépendre de ce qui se produit au niveau de la traduction.

Ce modèle peut expliquer un phénomène curieux. Dans certains cas, une mutation non-sens dans un des gènes d'une unité de transcription empêche l'expression des gènes suivants de cette unité. Cet effet est appelé **polarité** (voir Chapitre 8). Une conséquence fréquente est la disparition de l'ARNm correspondant à la partie suivante (distale) de l'unité de transcription.

Supposons qu'il puisse y avoir des terminateurs rho-dépendants à l'intérieur de l'unité de transcription, c'est-à-dire avant le terminateur qui est *couramment* utilisé ; les conséquences qui en découlent sont illustrées dans la **Figure 11.10**. Normalement ces terminateurs précoces ne sont pas utilisés, car les ribosomes empêchent rho de se fixer à l'ARN polymérase. Mais une mutation non-sens libère les ribosomes, si bien que le facteur rho devient libre de se déplacer sur l'ARNm, ce qui lui permet de se fixer à l'ARN polymérase pendant que l'enzyme est arrêtée au terminateur. Le résultat est une libération de l'enzyme, la conséquence étant que les régions distales de l'unité de transcription ne sont jamais exprimées. (Pourquoi y aurait-il des terminateurs internes ?). Peut-être correspondent-ils simplement à des épingles à cheveux qui miment le terminateur rho-dépendant, du fait d'une coïncidence fortuite de structures.

Le gène *rho* a été identifié par des mutations qui inhibent l'activité du facteur rho. En raison de la variété des effets de ces mutations plusieurs mutants furent

isolés comme ayant des phénotypes différents auxquels il fut donné des noms différents dont le plus connu est *suA* (reflétant son aptitude à supprimer la polarité).

Les mutations *rho* montrent de grandes variations dans leurs influences sur la terminaison. La nature première de l'effet est l'incapacité à terminer. Mais l'amplitude de l'effet dépend d'un locus cible déterminé, comme il a été vu dans le pourcentage de lecture *in vivo*. Ceci est probablement dû au caractère non absolu de la plupart des mutations *rho*, qui généralement permettent l'obtention d'une quantité d'activité relativement substantielle. Si les différents terminateurs ont besoin de quantités différentes de facteurs rho pour terminer, le taux d'activité résiduelle de rho peut conduire certains terminateurs à fonctionner (ceux actifs à basse concentration de rho), tandis que d'autres ne vont pas fonctionner, permettant au contraire à l'ARN polymérase de poursuivre la transcription.

L'aptitude des mutations *suA* à supprimer la polarité peut être expliquée si elles réduisent la probabilité que le facteur rho agisse sur un terminateur interne qui suit le codon non-sens. Dans ce cas, la terminaison de la traduction ne produit pas l'arrêt de la transcription ; et les régions de l'ARNm au-delà de la mutation peuvent être traduites par les ribosomes qui ré-initient plus loin.

Certaines mutations *rho* peuvent être supprimées par des mutations dans d'autres gènes. Cette approche fournit un excellent moyen pour identifier les protéines qui interagissent avec rho. La sous-unité β de l'ARN polymérase est impliquée dans deux types de mutations. Premièrement, les mutations dans le gène *rpoB* peuvent réduire la terminaison au site rho-dépendant. Deuxièmement les mutations dans *rpoB* peuvent restaurer la propriété de terminer la transcription aux sites rho-dépendants dans les mutants bactériens *rho*.

L'ANTITERMINAISON DÉPEND DE SITES SPÉCIFIQUES

La terminaison fournit une occasion pour certains phages de contrôler le passage des gènes précoces aux stades suivants de l'expression. La **Figure 11.11** compare l'utilisation de l'antiterminaison comme mécanisme de contrôle avec l'utilisation de nouveaux promoteurs.

Nous avons vu dans le Chapitre 9 que l'un des mécanismes de passage de la transcription des gènes précoces à l'expression de la phase suivante est le remplacement du facteur sigma de l'enzyme de l'hôte par un autre facteur qui va redicter ses spécificités d'initiation. Une alternative est de synthétiser une nouvelle ARN polymérase phagique. Dans l'un ou l'autre cas la caractéristique

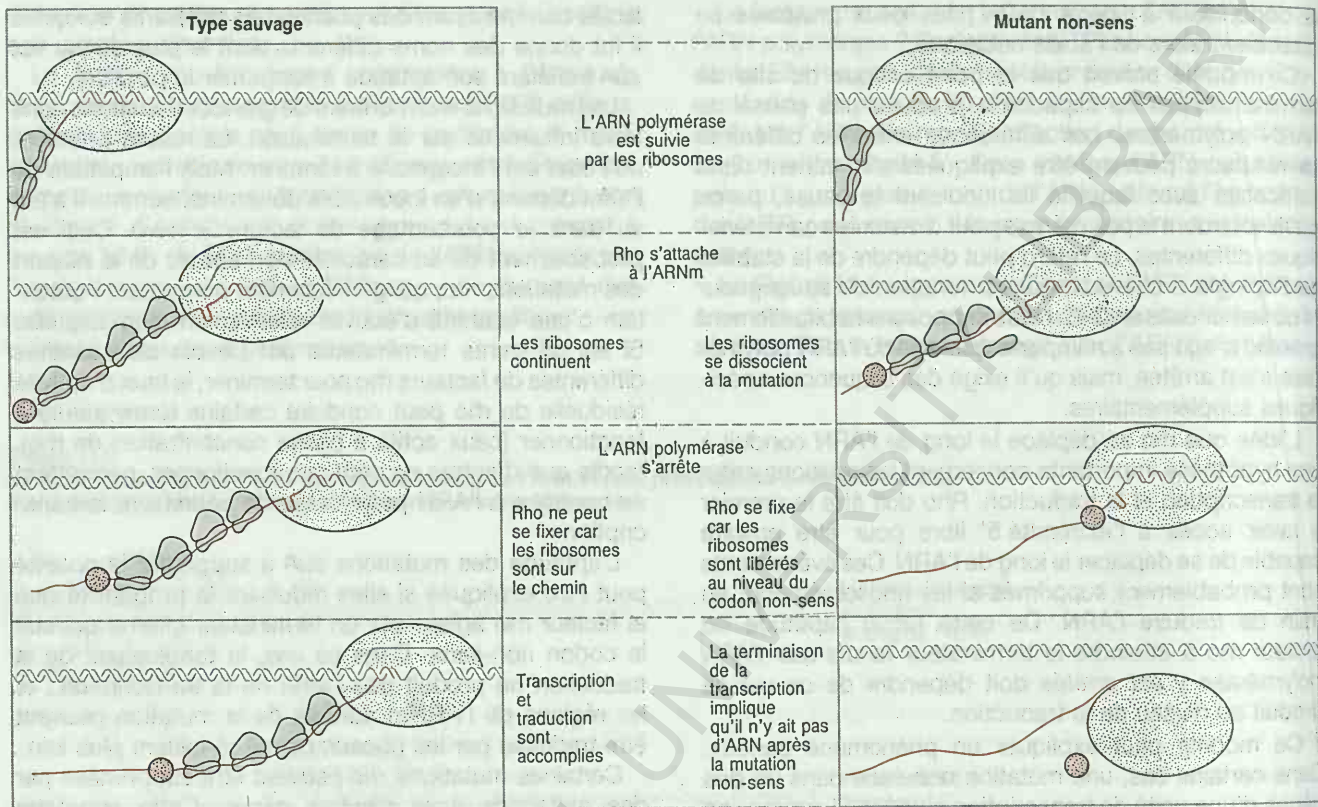


Figure 11.10

L'action du facteur rho peut créer un lien entre la transcription et la traduction quand le terminateur rho-dépendant se trouve peu après une mutation non-sens.

principale qui distingue le nouveau groupe de gènes tient au fait qu'ils contiennent des *promoteurs différents de ceux couramment reconnus par l'ARN polymérase de l'hôte*.

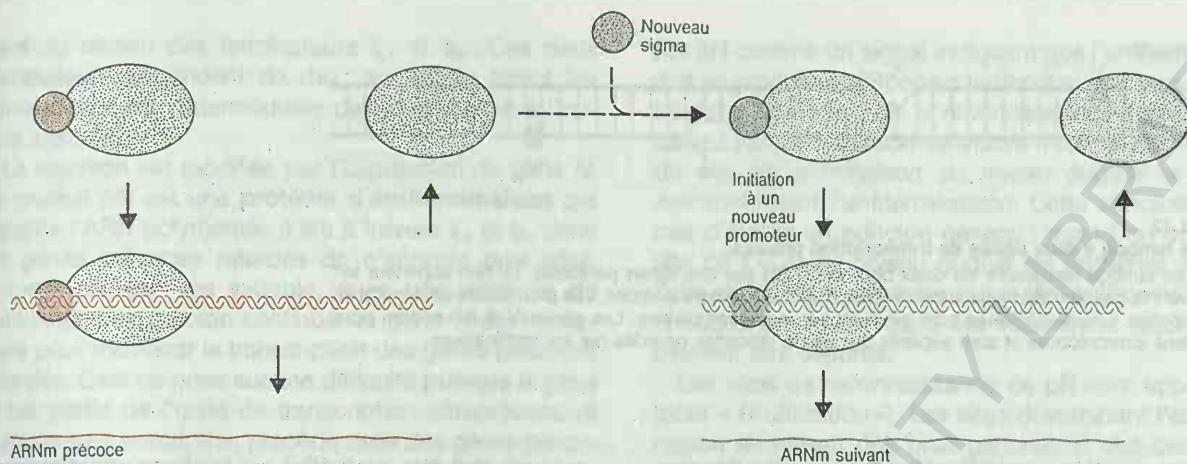
Dans ce cas l'expression d'une nouvelle série de gènes va dépendre de l'expression des gènes précoces après que le facteur rho ou la nouvelle polymérase ait été synthétisé. Les deux groupes de transcrits sont indépendants. L'expression des gènes précoces peut cesser quand le passage à l'étape suivante s'est produite.

En revanche l'utilisation d'antiterminateur dépend d'un arrangement particulier des gènes. Les gènes précoces se trouvent à côté des gènes qui vont être exprimés juste après, mais ils sont séparés d'eux par les sites de terminaison. Si la terminaison est empêchée au niveau de ces sites, l'ARN polymérase lira sans s'arrêter les gènes de l'autre côté. Dans l'antiterminaison le même promoteur continue donc à être reconnu par l'ARN polymérase. De cette façon les nouveaux gènes sont

exprimés uniquement par l'extension de la chaîne d'ARN, qui forme des molécules contenant les séquences des gènes précoces à l'extrémité 5' et les séquences des nouveaux gènes à l'extrémité 3'. Puisque les deux types de séquences restent liés, inévitablement l'expression des gènes précoces se poursuit.

L'exemple le plus caractéristique d'antiterminaison est fournie par le phage lambda. Le plan global de contrôle de l'ensemble est très similaire à celui trouvé dans d'autres phages : l'ARN polymérase de l'hôte transcrit les gènes précoces, et l'un des produits des gènes précoces est nécessaire à la transcription de la série suivante de gènes du phage. Dans le cas de lambda, les gènes dont l'initiation de la transcription se fait par l'ARN polymérase de l'hôte, sont appelés gènes **ultraprecoces** et ceux exprimés à l'étape suivante sont appelés gènes **precoces retardés**. La transition entre les deux étapes d'expression est contrôlée par la prévention de la terminaison à la fin des gènes ultraprecoces.

CONTRÔLE DE L'INITIATION



CONTRÔLE DE LA TERMINAISON

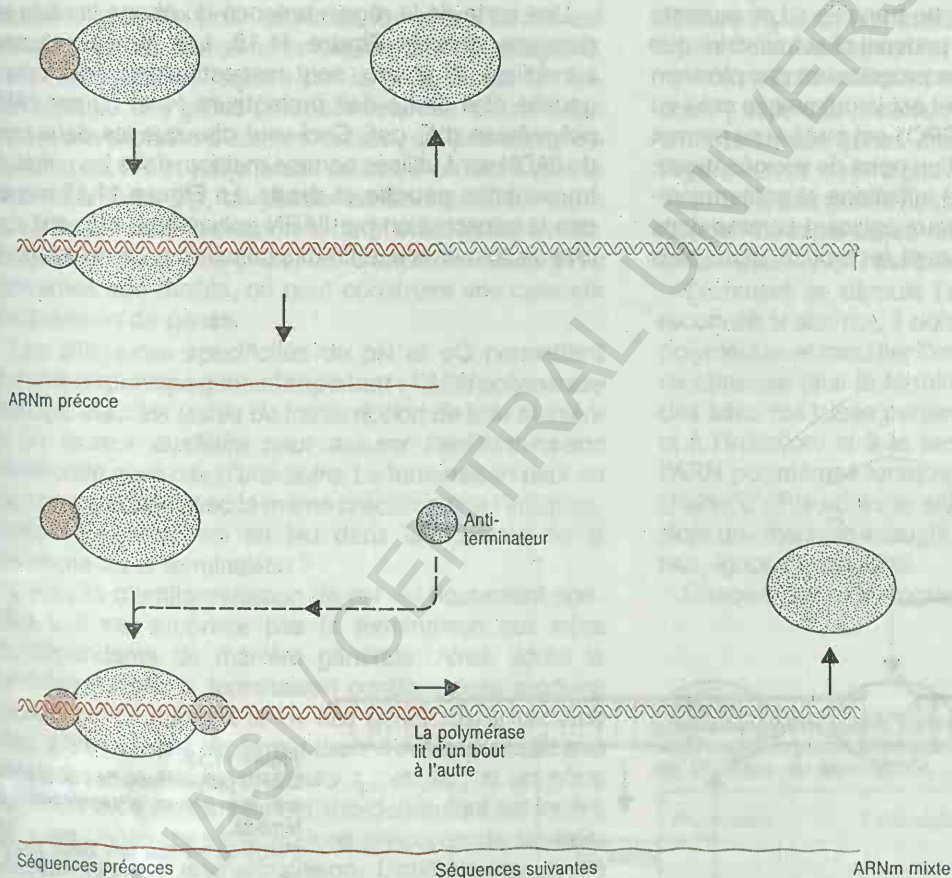


Figure 11.11
Les transitions dans la spécificité transcriptionnelle peuvent être contrôlées au niveau de l'initiation et de la terminaison.

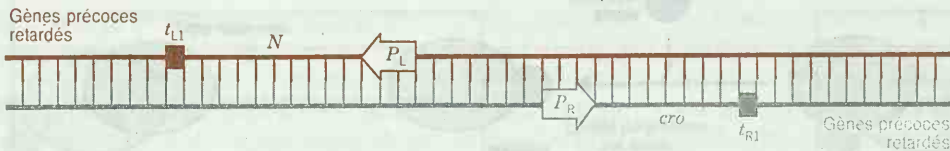


Figure 11.12

Le phage lambda a deux unités de transcription précoce.

La carte de lambda représente les deux brins de l'ADN par des lignes parallèles. Le brin supérieur est transcrit vers la gauche, alors que le brin inférieur est transcrit vers la droite. Les promoteurs sont indiqués par des flèches. Les terminateurs sont indiqués par les boîtes pleines. Les gènes *N* et *cro* codent pour les fonctions ultraprécoces et sont séparés des gènes précoces retardés par les terminateurs.

Le gène régulateur contrôlant le passage entre l'expression ultraprécoc et précoce retardée a été identifié par des mutations qui inhibent cette transition. Les mutants dans le gène *N* de lambda ne peuvent plus transcrire que les gènes ultraprécoces. Ils ne progressent pas plus loin dans le cycle infectieux. L'effet est le même que celui vu lorsque le gène 28 du phage SPO1 est muté et ne permet plus la production de σ_{gp28} . D'un point de vue génétique, les mécanismes de nouvelles initiations et antiterminaisons sont analogues. Tous deux agissent comme dans un **contrôle positif** dans lequel le produit d'un gène

précoce doit être synthétisé par le phage afin d'exprimer la série de gènes suivante.

Une carte de la région précoce du phage lambda est dessinée dans la Figure 11.12. Les gènes précoces immédiats, *N* et *cro*, sont respectivement transcrits à gauche et à droite des promoteurs P_L et P_R par l'ARN polymérase d'*E. coli*. Ceci veut dire que les deux brins de l'ADN sont utilisés comme matrice, dans les unités de transcription gauche et droite. La Figure 11.13 montre que la transcription par l'ARN polymérase d'*E. coli* s'arrête d'elle-même à la fin des gènes *N* et *cro*, respective-

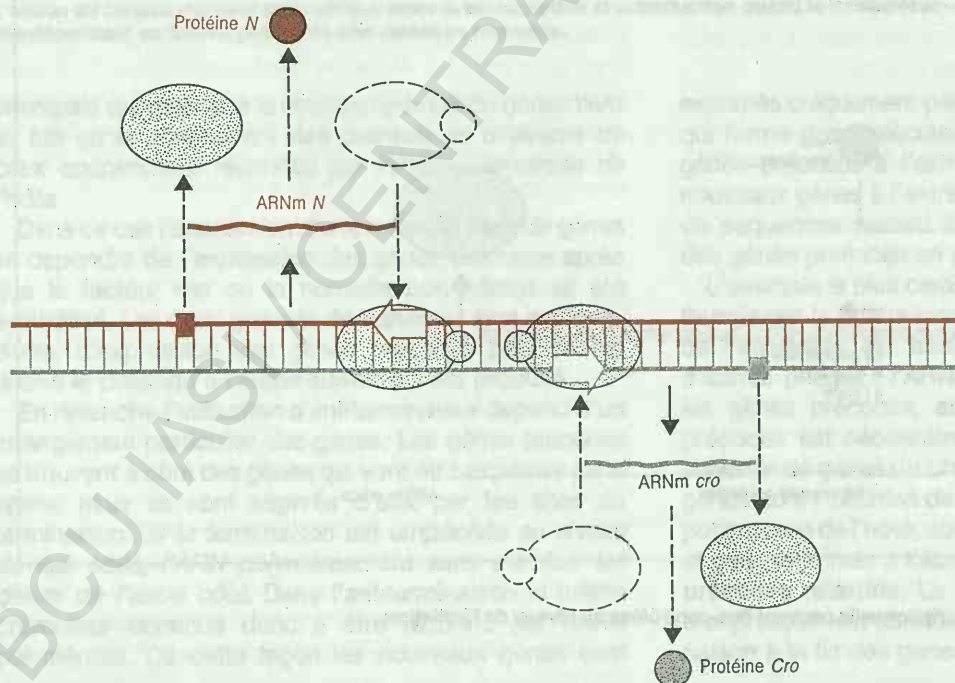


Figure 11.13

L'ARN polymérase de l'hôte transcrit les gènes ultraprécoces de lambda.

L'enzyme se fixe aux deux promoteurs P_L et P_R et transcrit les gènes *N* et *cro* respectivement jusqu'aux terminateurs t_{L1} et t_{R1} . La terminaison nécessite le facteur rho aux deux sites (ceci n'est pas montré dans la figure). La transcription de ces deux unités donne naissance à deux sortes d'ARNm ultraprécoces codant pour pN et Cro.

ment au niveau des terminateurs t_{L1} et t_{R1} . Ces deux terminateurs dépendent de ρ ; en fait ils furent les terminateurs par l'intermédiaire desquels ρ fut à l'origine identifié.

La situation est modifiée par l'expression du gène N . Le produit pN est une **protéine d'antiterminaison** qui entraîne l'ARN polymérase à lire à travers t_{L1} et t_{R1} dans les gènes précoces retardés de n'importe quel côté. Comme pN est très instable, avec une demi-vie de 5 minutes, l'expression continue du gène N est nécessaire pour maintenir la transcription des gènes précoces retardés. Ceci ne pose aucune difficulté puisque le gène N fait partie de l'unité de transcription ultraprécoc et puisque sa transcription précède celle des gènes précoces retardés.

Comme pour d'autres phages, un autre contrôle est encore nécessaire pour exprimer les gènes tardifs qui codent pour les composants de la particule phagique. Cette transition est régulée par le gène Q , qui fait lui-même partie des gènes précoces retardés. Son produit pQ est une protéine d'antiterminaison, qui permet spécifiquement à l'ARN polymérase initiant à un autre site, le promoteur tardif P_R , de passer au-delà du terminateur qui se trouve entre le promoteur tardif et les gènes tardifs. En utilisant donc des protéines d'antiterminaison avec différentes spécificités, on peut construire une cascade d'expression de gènes.

Les différentes spécificités de pN et pQ permettent d'établir un principe général important : *l'ARN polymérase interagit avec les unités de transcription de telle manière qu'un facteur auxiliaire peut assurer l'antiterminaison d'une unité mais pas d'une autre*. La terminaison peut en fait être contrôlée avec la même précision que l'initiation. Quels sites sont mis en jeu dans le contrôle de la spécificité de la terminaison ?

L'activité d'antiterminaison de pN est hautement spécifique. Il ne supprime pas la terminaison aux sites ρ -dépendants de manière générale. Ainsi, après la synthèse de pN, la terminaison continue à se produire comme de coutume à la fin des gènes de la bactérie hôte. *L'événement de terminaison n'est cependant pas déterminé par les terminateurs t_{L1} et t_{R1} ; si un gène bactérien avec un terminateur ρ -dépendant est inséré dans les unités de transcriptions précoces de lambda, pN produira son antiterminaison. L'antiterminaison se produit ainsi à n'importe quel terminateur rencontré par l'ARN polymérase qui a traversé l'unité de transcription précoce de lambda. Ces interactions sont résumées dans le Tableau 11.1.*

L'une et l'autre des unités de transcriptions précoces de lambda doivent posséder des sites qui sont reconnus

par pN comme un signal indiquant que l'antiterminaison doit se produire au prochain terminateur qui sera rencontré. En d'autres termes, la *reconnaissance du site nécessaire à l'antiterminaison se trouve à une place différente du site de terminaison au niveau duquel se produit éventuellement l'antiterminaison*. Cette conclusion permet d'établir un principe général : quand on connaît le site de l'ADN au niveau duquel les protéines exercent leurs effets, nous ne pouvons pas dire qu'il coïncide avec la séquence d'ADN qui a été initialement reconnue. Ils peuvent être séparés.

Les sites de reconnaissance de pN sont appelés *nut* (pour « *N utilization* »). Les sites déterminant l'antiterminaison au niveau des bras gauches et des bras droits sont désignés respectivement par *nutL* et *nutR*, et nous savons qu'ils doivent se trouver entre P_L et t_{L1} d'un côté, et entre P_R et t_{R1} de l'autre côté. Où sont-ils exactement situés ? La cartographie des mutations *nut⁻* et des mutations ponctuelles localise *nutL* entre le point d'initiation P_L et le début de la région codant pour N , et *nutR* entre la fin des gènes *cro* et t_{L1} . Comme le présente la **Figure 11.14** ceci signifie que les deux sites *nut* se trouvent à des positions différentes par rapport à l'organisation de leurs unités de transcription. Alors que *nutL* est au début, *nutR* est près du terminateur.

Comment se déroule l'antiterminaison ? Lorsque pN reconnaît le site *nut*, il doit interagir à ce site avec l'ARN polymérase et modifier l'enzyme de telle sorte qu'elle ne reconnaisse plus le terminateur. La localisation variable des sites *nut* laisse penser que cet événement n'est lié ni à l'initiation, ni à la terminaison, mais peut arriver à l'ARN polymérase lorsqu'elle est en train d'allonger une chaîne d'ARN après le site *nut*. La polymérase devient alors une machine aveugle qui continue après le terminateur, ignorant ce signal.

Chaque site *nut* contient une séquence de 17 pb,

Tableau 11.1
L'antiterminaison atteint toutes les ARN polymérases qui initient au niveau des promoteurs précoces de lambda, indépendamment de la nature du terminateur.

Promoteur Terminateur		Que se produit-il dans les cellules infectées par lambda ?
Bactérien	Bactérien	La terminaison se produit
Lambda précoce	Lambda précoce	pN prévient la terminaison
Lambda précoce	Bactérien	pN prévient la terminaison

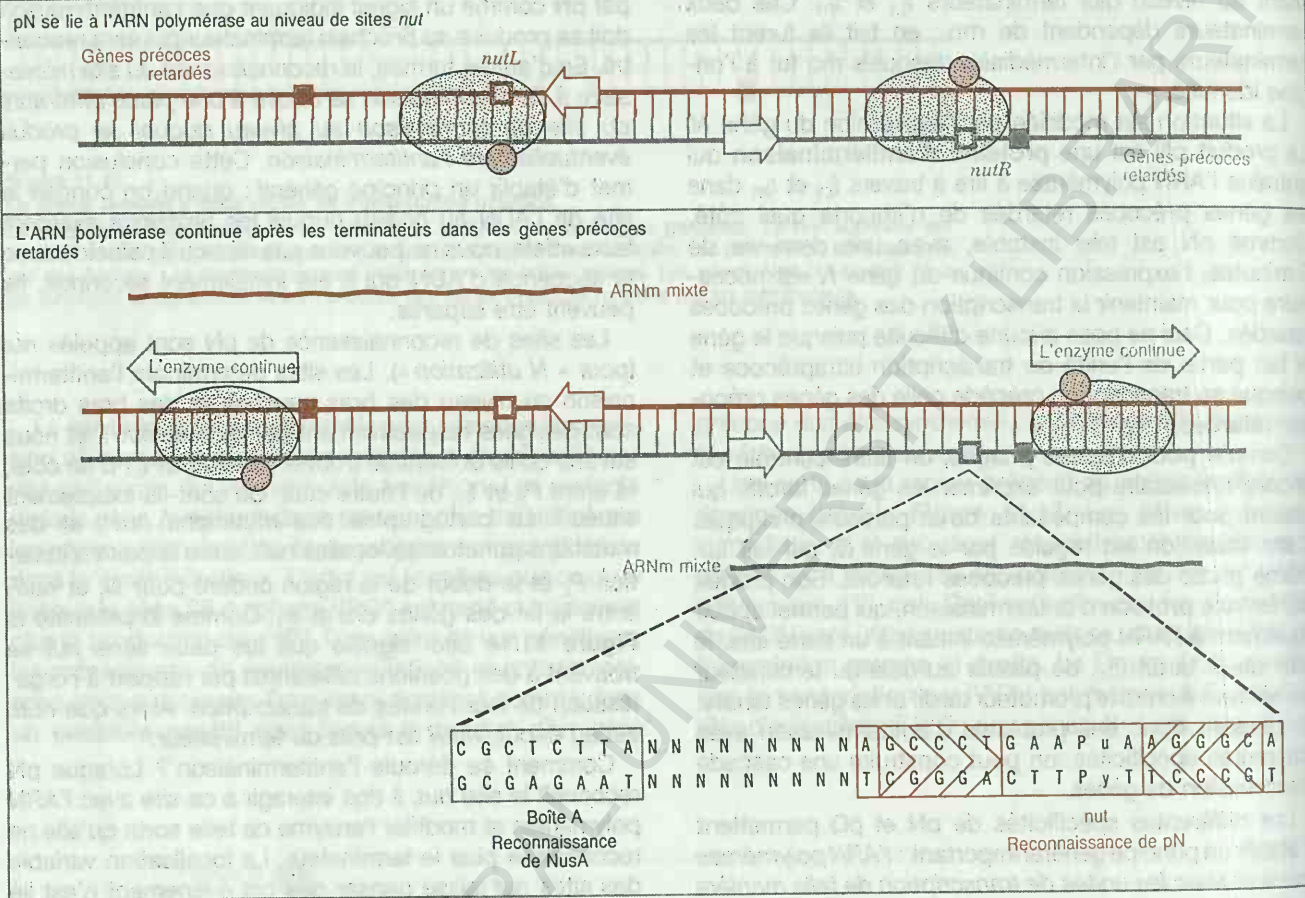


Figure 11.14

La protéine pN se fixe à l'ARN polymérase lorsqu'elle rencontre un site *nut* ; sa présence conduit l'enzyme à lire au-delà des terminateurs t_L et t_R , produisant ainsi un ARNm mixte qui contient à la fois les séquences des gènes précoces retardés et ultraprécoces.

identique à l'exception d'une seule position, orientée dans le même sens par rapport à la transcription. Ces séquences contiennent une répétition inversée de 5 pb séparée par une région non symétrique. Puisque des mutations *nut* ont été trouvées en dehors de cette région, on pense que les séquences reconnues par pN peuvent s'étendre au-delà de cette zone centrale.

La propriété de pN de reconnaître une courte séquence à l'intérieur de l'unité de transcription fournit-elle un

exemple d'un mécanisme d'antitermination plus largement utilisé ? D'autres phages apparentés à lambda ont des gènes *N* différents et différentes spécificités d'antitermination. La région du génome phagique dans laquelle se trouvent les sites *nut* a une séquence différente dans chacun de ces phages, et probablement chaque phage a ses sites *nut* caractéristiques reconnus spécifiquement par ses propres pN. Chacun de ces produits pN doit avoir la même capacité globale d'interagir avec l'appareil de

transcription dans le cadre de l'antiterminaison, mais a des spécificités différentes pour la séquence d'ADN qui active le mécanisme.

AVANTAGE DE SOUS-UNITÉS POUR L'ARN POLYMÉRASE ?

La découverte de l'antiterminaison comme un mécanisme de contrôle du phage a conduit à l'identification d'autres composants de l'appareil de transcription. Les protéines bactériennes avec lesquelles pN interagit peuvent être identifiées en isolant des mutants de *E. coli* dans lesquels pN est inefficace. Ces mutants ne peuvent pas être infectés efficacement par lambda parce que le phage est limité à l'expression de ses seuls gènes ultraprécoces. Plusieurs de ces mutations résident dans le gène *rpoB*. Ceci indique que pN (comme le facteur rho) interagit avec la sous-unité β de l'enzyme cœur.

Les autres mutants de *E. coli* qui empêchent la fonction de pN permettent d'identifier les loci *nus*, qui sont au moins au nombre de trois : *nusA*, *nusB* et *nusE* (le terme « *nus* » est un acronyme de l'anglais « *N utilization substance* »). Les loci *nus* doivent coder pour des protéines qui forment une partie de l'appareil de transcription, mais qui ne sont pas isolées avec l'ARN polymérase dans sa forme habituelle. Les fonctions *nusA* et *nusB* semblent intervenir seulement dans la terminaison de la transcription. *NusE* code pour la protéine ribosomale S10 ; la relation entre sa localisation dans la sous-unité 30S et sa fonction dans la terminaison n'est pas claire.

La protéine NusA est étroitement liée à l'action de pN ; on peut obtenir des mutations dans *N* qui corrigent l'obstacle à l'antiterminaison imposée par les mutations *nusA*. L'idée que NusA est nécessaire à l'action de pN est confortée par la capacité des deux protéines à se lier entre elles *in vitro*. Le polypeptide pN est petit (13 500 daltons), très basique, et de forme plutôt asymétrique. *In vitro*, le produit NusA de 69 000 daltons et le produit NusB de 15 000 daltons sont nécessaires en quantité équimolaire pour l'antiterminaison.

NusA fonctionne *in vitro* comme un *facteur de terminaison* aux sites t_{L2} et t_{R2} , qui sont classés comme sites rho-indépendants, et résident dans le génome à distance des sites rho-dépendants t_{L1} et t_{R1} . NusA permet l'arrêt de l'ARN polymérase à ces terminateurs pendant un long moment, 15 minutes environ. La libération de la chaîne d'ARN à l'état naissant nécessite la présence du facteur rho, mais à très faible concentration (qui est inefficace par

elle-même). Ceci suggère que l'action de NusA n'est pas la même que celle de rho, mais lui est complémentaire.

L'interaction entre NusA et pN peut dépendre des séquences au voisinage des sites *nus*. Juste en amont de chacun des sites *nus* se trouve une courte séquence « consensus » appelée *boîte A* (voir la Figure 11.14) qui peut être requise pour la reconnaissance de NusA. Dans ce cas, NusA et pN peuvent reconnaître des séquences adjacentes ou partiellement chevauchantes.

La protéine NusA se lie à l'enzyme « cœur » de la polymérase mais ne se lie pas à l'holoenzyme. Quand le facteur σ est ajouté au complexe $\alpha_2\beta\beta'$ NusA, il déplace la protéine NusA, reconstituant ainsi l'holoenzyme $\alpha_2\beta\beta'\sigma$. Ceci suggère que l'ARN polymérase passe par le cycle illustré dans la Figure 11.15 dans lequel elle existe sous la forme d'une enzyme prête à initier ($\alpha_2\beta\beta'\sigma$) et d'une enzyme prête à terminer.

Quand l'holoenzyme ($\alpha_2\beta\beta'\sigma$) se lie à un promoteur, elle libère son facteur σ , et produit ainsi l'enzyme cœur ($\alpha_2\beta\beta'$), qui synthétise l'ARN. Puis une protéine NusA reconnaît l'enzyme cœur et se fixe à elle, produisant le complexe $\alpha_2\beta\beta'$ NusA. Nous ne savons pas à quel stade cela arrive ; cela pourrait être à n'importe quel moment entre l'initiation et l'élongation. Quoi qu'il en soit, alors que la polymérase $\alpha_2\beta\beta'$ NusA est liée à l'ADN, le constituant NusA ne peut pas être déplacé. Mais quand la terminaison survient, l'enzyme est relâchée dans un état dans lequel NusA est automatiquement libéré ou bien peut être déplacé par le facteur σ .

Du point de vue de l'enzyme cœur, la vie consiste en une succession d'associations avec σ pour l'initiation et d'associations avec NusA et NusB pour la terminaison. σ et les facteurs Nus sont des associés mutuellement incompatibles du cœur. Il ne semble pas y avoir de raisons pour considérer davantage l'un d'eux plutôt que les deux autres comme un élément du complexe, mais il faut les considérer comme des sous-unités alternatives. Ainsi l'enzyme cœur représente une forme minimale de l'ARN polymérase compétente pour s'engager dans les fonctions de base de la synthèse d'ARN mais manquant des sous-unités nécessaires pour les autres fonctions. On pourrait discuter à l'infini de la question de savoir où finit l'ARN polymérase et où commence, au sens plus large, l'appareil de transcription.

Le composant NusA peut servir de lien entre pN et l'enzyme cœur. Cela a un sens si NusA est une sous-unité de terminaison, dont l'incorporation dans l'enzyme confère une spécificité vis-à-vis de certains terminateurs. Ainsi pN peut agir comme antitermineur en empêchant NusA d'exercer sa fonction. Cela soulève immédiatement

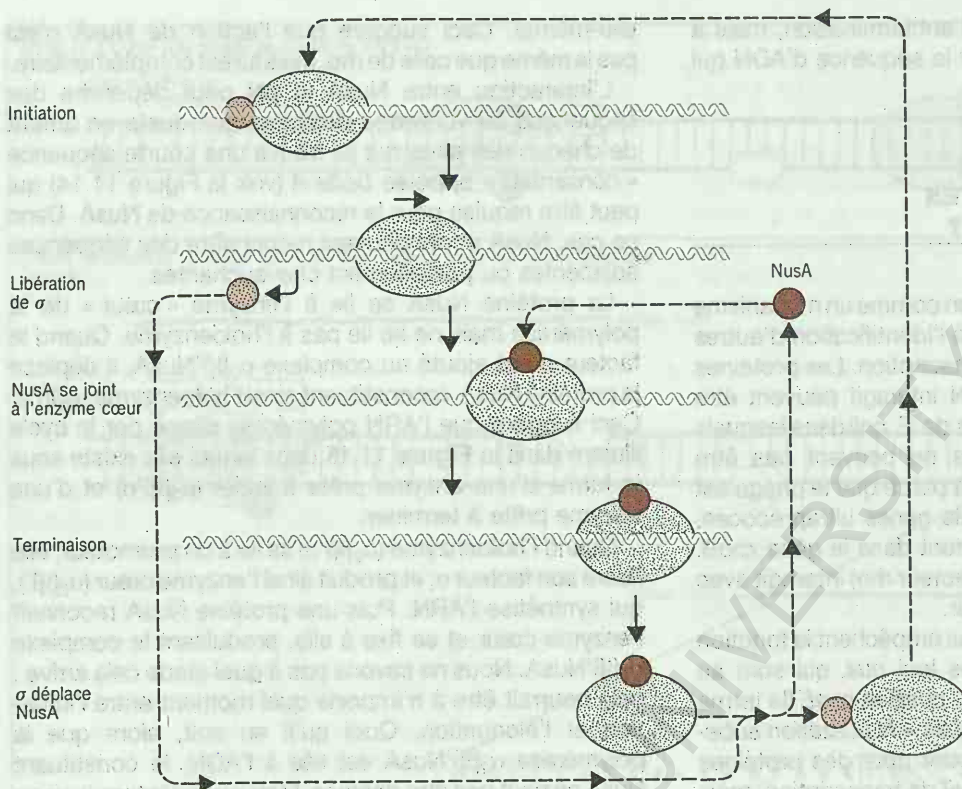


Figure 11.15

L'ARN polymérase peut alterner entre une forme compétente pour l'initiation et une forme compétente pour la terminaison selon que sigma ou NusA se remplacent l'un l'autre sur l'enzyme cœur.

la possibilité que la bactérie hôte puisse contenir d'autres protéines, analogues à NusA (ou à pN) dont les interactions avec l'enzyme cœur contrôlraient la reconnaissance des sites de terminaisons.

Nous ne savons pas exactement comment l'enzyme cœur, les protéines Nus et le facteur rho coopèrent dans la terminaison, ni comment relier clairement leurs activités à la classification des terminateurs en sites rho-dépendants et rho-indépendants. L'existence de NusA implique que l'enzyme cœur ne peut jamais effectuer la terminaison toute seule ; ainsi les expériences *in vitro* avec de l'enzyme cœur purifiée pourrait refléter une activité aberrante. Des sites rho-dépendants et indépendants sont reconnus dans des conditions quelque peu différentes *in vitro* (force ionique basse plutôt que haute) ; ce qui a toujours soulevé l'inquiétude qu'un tel ensemble de conditions pouvaient ne pas refléter la situation naturelle. L'utilisation de ces types de sites *t* nécessite d'être réexaminée avec des

systèmes *in vitro* qui contiennent les autres constituants potentiels de la terminaison.

POUR EN SAVOIR PLUS

Les mécanismes et la régulation de la terminaison bactérienne ont été passés en revue par Adhya & Gottesman (*Ann. Rev. Biochem.* 47, 967-996, 1978) et Platt (*Ann. Rev. Biochem.* 55, 339-372, 1986). L'atténuation a été passée en revue par Bauer et al. dans *Gene Function in Prokaryotes* (pp. 65-89). La nature des signaux de terminaisons a été passée en revue dans le contexte des séquences consensus par Rosenberg & Court (*Ann. Rev. Genet.* 13, 319-353, 1979). La découverte de l'appareil de transcription bactérien élargi a été rassemblée par Greenblatt (*Cell* 24, 8-9, 1981). La terminaison eucaryote a été passée en revue par Birnstiel (*Cell* 41, 349-359, 1985). L'atténuation de l'opéron *trp* a été mise à jour par Yanofsky (*Nature* 289, 751-758, 1981) et Platt (*Cell* 24, 10-23, 1981). Les fonctions de cet opéron ont été résumées en terme de séquence par Yanofsky et al. (*Nuc. Acids Res.* 9, 6647-6668, 1981).

CHAPITRE 12

LES CASCADES LYTIQUES ET LA RÉPRESSION LYSOGÉNIQUE

Certains phages n'ont qu'une seule stratégie pour survivre. En infectant un hôte sensible, ils détournent ses fonctions dans le but de produire un grand nombre de nouvelles particules virales. Le résultat de cette **infection lytique** est la mort de la bactérie. Dans un cycle lytique typique, l'ADN (ou l'ARN) du phage entre dans la bactérie hôte ; ces gènes sont transcrits en ordre régulier, le matériel génétique du phage est répliqué et les protéines qui constituent la particule virale sont synthétisées. Finalement, la bactérie hôte est détruite (**lyse**) libérant les particules phagiques nouvellement synthétisées.

D'autres phages peuvent avoir un destin double. D'une part, ils sont capables de se perpétuer dans un cycle lytique, utilisant une stratégie destinée à produire autant de copies du phage aussi rapidement que possible. D'autre part, comme alternative, ils peuvent avoir une existence « masquée » dans laquelle le génome est présent dans la bactérie sous une forme latente appelée **prophage**. Cette forme de propagation est appelée **lysogénie**.

Dans les bactéries lysogéniques, le prophage est **intégré** dans le génome bactérien, et il est transmis de la même façon que les gènes bactériens. Une bactérie

lysogène acquiert, grâce à la présence du prophage, l'**immunité** contre d'ultérieures infections par des particules phagiques du même type ; c'est ainsi qu'il n'y a généralement qu'une seule copie de prophage par génome bactérien.

Des transitions peuvent se produire entre les modes d'existence lysogénique et lytique, comme l'illustre la **Figure 12.1**. Quand un phage produit d'un cycle lytique entre dans une nouvelle cellule bactérienne hôte, il peut soit reproduire un cycle lytique, soit entrer dans l'état lysogénique. Le choix dépend des conditions de l'infection et du génotype du phage et de la bactérie. Un prophage peut être libéré de la restriction lysogénique par un mécanisme appelé **induction** au cours duquel il est **excisé** du génome bactérien, engendrant un ADN libre qui peut s'engager dans un cycle lytique.

Un autre type d'existence dans les bactéries est représenté par les **plasmides**. Ce sont des éléments autonomes dont les génomes existent dans la cellule sous forme **extrachromosomique**. Les plasmides sont des molécules d'ADN circulaires autorépliquatives qui sont maintenues dans la cellule en un nombre de copies stables et caractéristiques. Le nombre reste constant de génération en génération.

Infection

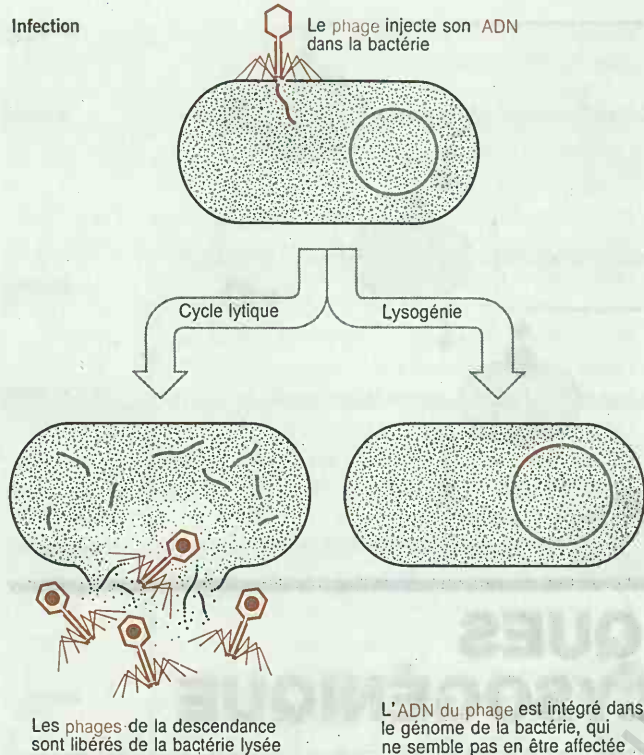


Figure 12.1

Le développement lytique implique la production de particules phagiques avec destruction de la bactérie hôte, tandis que l'existence lysogénique permet au génome du phage d'être porté comme une partie de l'information génétique de la bactérie.

Certains de ces éléments ont aussi un mode de vie alternatif. Ils peuvent exister, soit sous une forme extra-chromosomique autonome, soit être insérés dans le chromosome bactérien, étant alors transmis comme une partie de son propre génome. Ils sont, dans ce cas, correctement appelés **épisomes** (mais parfois les termes de plasmide et d'épisome sont utilisés improprement comme s'ils étaient interchangeables).

Comme les phages lysogéniques, les plasmides et les épisomes gardent égoïstement possession de leur bactérie, rendant souvent impossible l'établissement d'un autre élément du même type. Cet effet est aussi appelé **immunité**, quoique la base de l'immunité plasmidique soit différente de celle de l'immunité lysogénique (voir le Chapitre 13).

LE DÉVELOPPEMENT LYTIQUE EST CONTRÔLÉ PAR UNE CASCADE

Le génome des phages est, par nécessité, petit. En effet, comme dans tous les virus, la principale contrainte réside dans la difficulté à encapsider l'acide nucléique dans l'enveloppe protéique. Cette limitation impose en grande partie les stratégies virales de reproduction. Généralement, le virus détourne la machinerie de la cellule hôte qui, au lieu d'exprimer et de répliquer les gènes bactériens, réplique et exprime les gènes phagiques.

Habituellement, les gènes du phage gouvernent des fonctions qui garantissent une réplication préférentielle de l'ADN phagique. Ces fonctions peuvent être en rapport avec l'initiation de la réplication ou peuvent fournir une nouvelle ADN polymérase. Mais des changements sont toujours introduits dans la capacité de la cellule hôte à initier la transcription. Elles peuvent entraîner le remplacement de l'ARN polymérase ou modifier ses propriétés d'initiation et de terminaison. Le résultat est toujours le même : les ARNm du phage sont transcrits préférentiellement. Puis, pour ce qui est des synthèses protéiques, le phage se contente habituellement d'utiliser la machinerie de l'hôte, reprogrammant ses activités principalement grâce au remplacement des ARN bactériens par les ARN du phage.

Le développement lytique est accompli selon un schéma dans lequel les fonctions du phage sont exprimées dans un ordre déterminé. Ceci garantit la présence d'une concentration correcte de chaque constituant au moment approprié. Le cycle peut être divisé en deux grandes parties, comme c'est illustré dans la Figure 12.2.

L'**infection précoce** décrit la période qui va de l'entrée de l'ADN au début de sa réplication. L'**infection tardive** définit la période s'étendant du début de la réplication à l'étape finale de lyse des bactéries, lyse par laquelle sont libérées les particules phagiques de la descendance. Dans l'ordre habituel de bataille, la phase précoce est dévolue à la production des enzymes de réplication de l'ADN. Ceci inclut les enzymes impliquées dans la synthèse, la recombinaison et parfois la modification de l'ADN. Leurs activités provoquent l'accumulation de **pool** de génome phagique. Dans ce pool, les génomes sont sans cesse répliqués et recombinaisonnés, de telle sorte que les événements survenant pendant un seul cycle lytique concernent toute une population de génomes phagiques.

Quand la phase tardive commence, les protéines constituant la particule du phage sont synthétisées. De très nombreuses protéines différentes sont souvent nécessaires pour faire les structures de la tête et de la queue ;

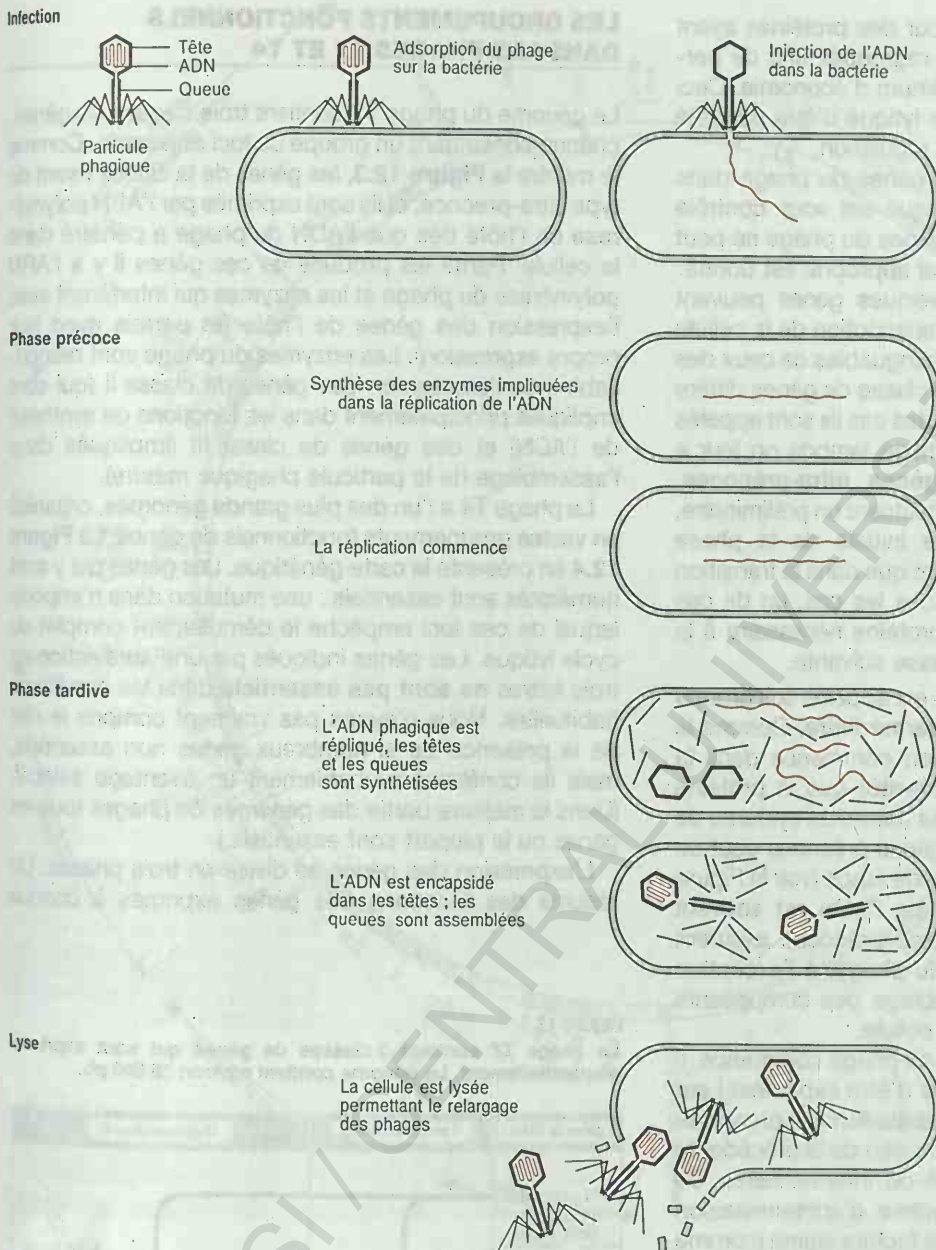


Figure 12.2

Le développement lytique se met en place en produisant des génomes de phage et les particules protéiques qui sont assemblées pour produire les phages de la descendance.

Les particularités de ce cycle sont à comparer avec l'ensemble du processus révélé par la microscopie électronique des bactéries infectées (Figure 1.1).

ainsi, la plus grande partie du génome du phage contient des fonctions tardives. En plus des protéines de structure, des protéines d'assemblage sont nécessaires pour réaliser la construction de la particule mais n'y sont pas elles-mêmes incorporées. Au moment où ces composants sont assemblés dans la tête et dans la queue, la

réplication de l'ADN a atteint son niveau maximum. Les génomes sont alors insérés dans les têtes protéiques vides, les queues sont ajoutées, et la cellule est lysée.

L'organisation de la carte génétique du phage reflète souvent le déroulement du développement lytique. Le concept d'opéron peut être considéré comme un extrême

dans lequel les gènes codant pour des protéines ayant des fonctions apparentées sont regroupés afin de permettre leur contrôle avec le maximum d'économie. Ceci permet au déroulement du cycle lytique d'être contrôlé par un petit nombre de voies de régulation.

Pour assurer l'expression des gènes du phage dans un ordre déterminé, le cycle lytique est sous contrôle positif. Chaque groupement de gènes du phage ne peut être exprimé que lorsque le signal approprié est donné.

Habituellement, seulement quelques gènes peuvent être exprimés par l'appareil de transcription de la cellule hôte. Leurs promoteurs sont indistinguables de ceux des gènes de l'hôte. Le nom de cette classe de gènes diffère selon les phages. Dans la plupart des cas ils sont appelés les **gènes précoces**. Dans le phage lambda on leur a donné le nom évocateur de **gènes ultra-précoces**. Comme leur nom l'indique, ils constituent un préliminaire, correspondant juste à la partie initiale de la phase précoce. Parfois, ils n'interviennent que dans la transition vers la période suivante. Dans tous les cas, un de ces gènes code toujours pour une protéine nécessaire à la transcription des gènes de la classe suivante.

Cette seconde classe de gènes est appelée soit groupe **précoce retardé**, soit groupe **intermédiaire**. Comme le suggère leur nom, leur expression commence dans la phase précoce, généralement aussitôt que la protéine régulatrice est disponible. Selon la nature du système de régulation, la première série de gènes précoces peut ou non continuer à être exprimée à cette étape (voir la Figure 11.11). L'expression des gènes de l'hôte est souvent réduite. Les deux séries de gènes précoces assurent ensemble toutes les fonctions du phage, à l'exception de celles nécessaires à l'assemblage des composants de l'enveloppe et à la lyse de la cellule.

Quand la réplication de l'ADN du phage commence, il est temps pour les **gènes tardifs** d'être exprimés. Leur transcription à cette étape est habituellement organisée par un nouveau gène de régulation issu de la précédente série de gènes (précoce retardé ou intermédiaire). Ce régulateur peut être un autre facteur d'antitermination (comme dans lambda) ou un autre facteur sigma (comme dans SPO1).

L'utilisation de ces contrôles successifs, dans lesquels chaque série de gènes contient un régulateur qui est nécessaire à l'expression de la série suivante, crée une cascade dans laquelle les groupes de gènes sont activés et parfois inactivés à des moments précis. La construction détaillée de chaque cascade est différente selon le phage. Mais les résultats sont identiques, comme le montre la partie suivante.

LES GROUPEMENTS FONCTIONNELS DANS LES PHAGES T7 ET T4

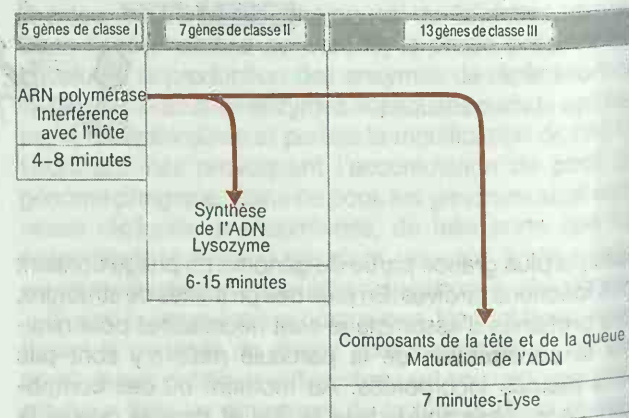
Le génome du phage T7 contient trois classes de gènes, chacun constituant un groupe de loci adjacents. Comme le montre la **Figure 12.3**, les gènes de la classe I sont du type ultra-précoce, et ils sont exprimés par l'ARN polymérase de l'hôte dès que l'ADN du phage a pénétré dans la cellule. Parmi les produits de ces gènes il y a l'ARN polymérase du phage et les enzymes qui interfèrent avec l'expression des gènes de l'hôte (et parfois avec leur propre expression). Les enzymes du phage sont responsables de l'expression des gènes de classe II (qui sont impliqués principalement dans les fonctions de synthèse de l'ADN) et des gènes de classe III (impliqués dans l'assemblage de la particule phagique mature).

Le phage T4 a l'un des plus grands génomes, organisé en vastes groupements fonctionnels de gènes. La **Figure 12.4** en présente la carte génétique. Les gènes qui y sont numérotés sont essentiels : une mutation dans n'importe lequel de ces loci empêche le déroulement complet du cycle lytique. Les gènes indiqués par une abréviation de trois lettres **ne sont pas essentiels** dans les conditions habituelles. Nous n'avons pas vraiment compris le rôle de la présence de si nombreux gènes non essentiels, mais ils confèrent probablement un avantage sélectif. (Dans la majeure partie des génomes de phages tous les gènes ou la plupart sont essentiels.)

L'expression des gènes se divise en trois phases. Un résumé des fonctions des gènes exprimés à chaque

Figure 12.3

Le phage T7 contient 3 classes de gènes qui sont exprimés séquentiellement. Le génome contient environ 38 000 pb.



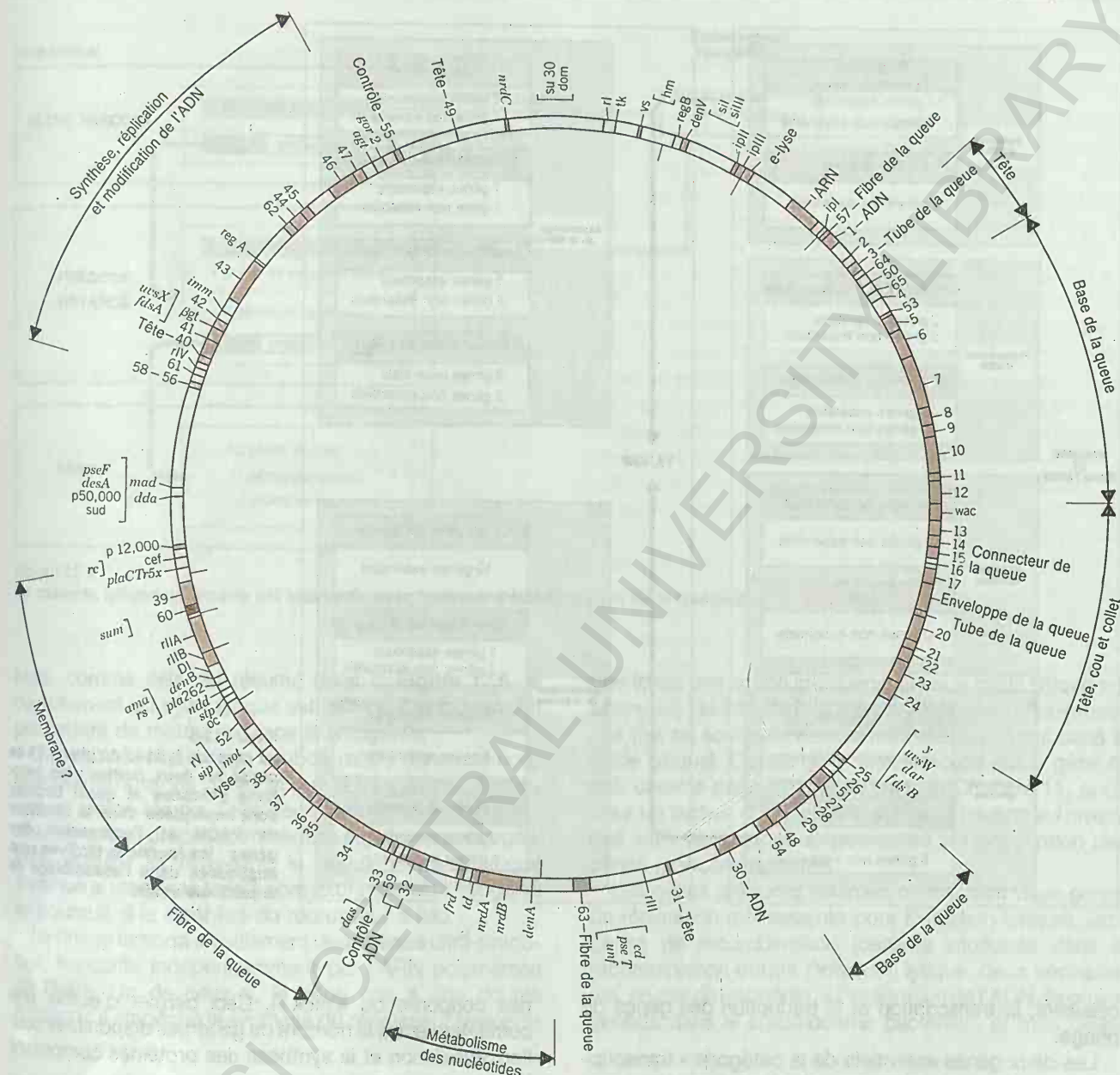


Figure 12.4
La carte du phage T4 est circulaire et montre des groupements de fonctions apparentées dans le génome de 165 000 pb.

étape est donné dans la **Figure 12.5**. La transcription des gènes précoces est réalisée par l'ARN polymérase de l'hôte. Légèrement plus tard, les gènes quasi tardifs sont transcrits, mais nous n'avons pas encore compris comment ils étaient contrôlés (leurs noms reflètent un

mode d'expression qui ne correspond pas avec les catégories que nous avons décrites pour d'autres phages). Les gènes précoces et quasi tardifs rendent compte de presque toutes les fonctions du phage concernant la synthèse de l'ADN, les modifications de la structure

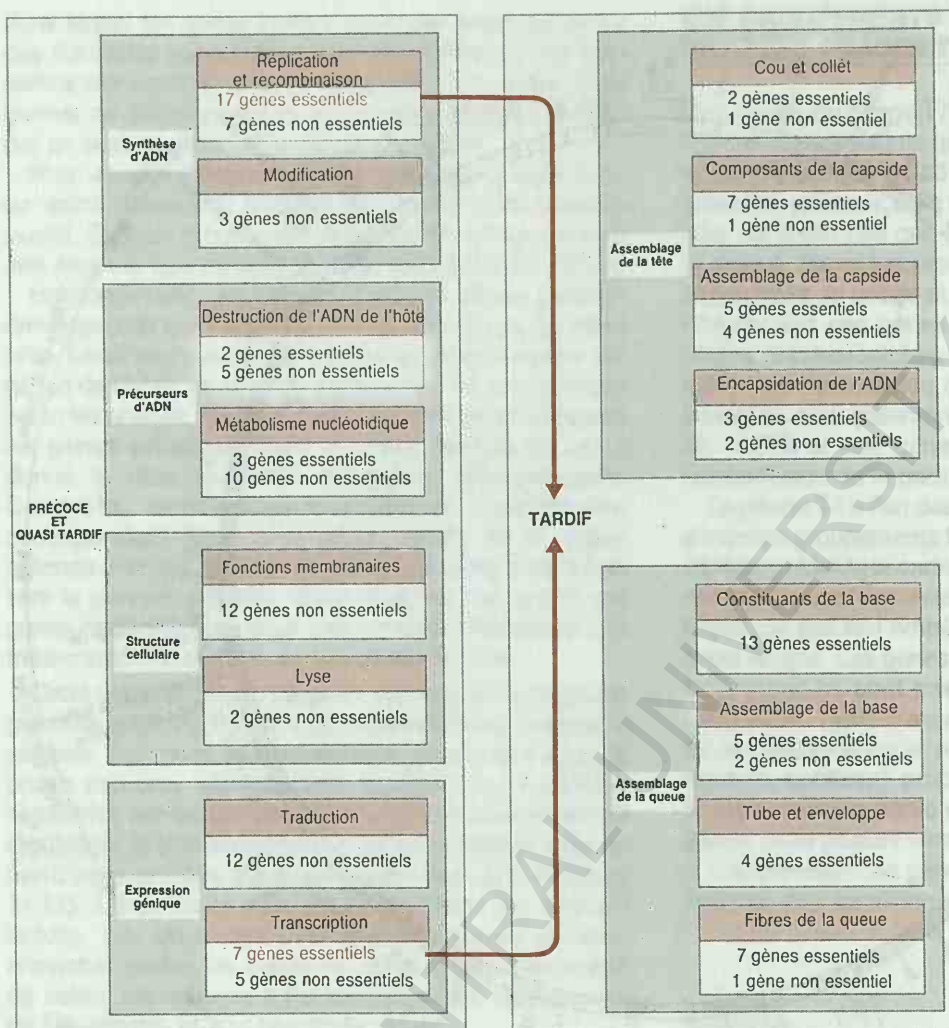


Figure 12.5

La cascade lytique du phage T4 se divise en deux parties : les fonctions précoces et quasi tardives sont impliquées dans la synthèse de l'ADN et l'expression des gènes ; les fonctions tardives sont impliquées dans l'assemblage de la particule virale.

cellulaire, la transcription et la traduction des gènes du phage.

Les deux gènes essentiels de la catégorie « transcription » remplissent une fonction de régulation : leurs produits sont nécessaires à l'expression des gènes tardifs. L'infection par le phage T4 dépend d'un lien quasi mécanique entre la réplication et l'expression des gènes tardifs. En effet, seul l'ADN en réplication active peut être utilisé comme matrice pour la transcription des gènes tardifs. Cette relation est engendrée par l'introduction de nouveaux facteurs sigma et par d'autres modifications de l'ARN polymérase de l'hôte la rendant dépendante de caractéristiques de l'ADN en réplication (probablement

des coupures ou « nick »). Ceci permet d'établir une corrélation entre le nombre de génomes disponibles pour l'encapsidation et la synthèse des protéines composant le phage.

LA CASCADE LYTIQUE DE LAMBDA REPOSE SUR L'ANTITERMINAISON

Un des systèmes de cascade le plus complexe est fourni par le phage lambda. La cascade du développement lytique par elle-même est directe, avec deux régulateurs qui contrôlent les étapes successives du développement.

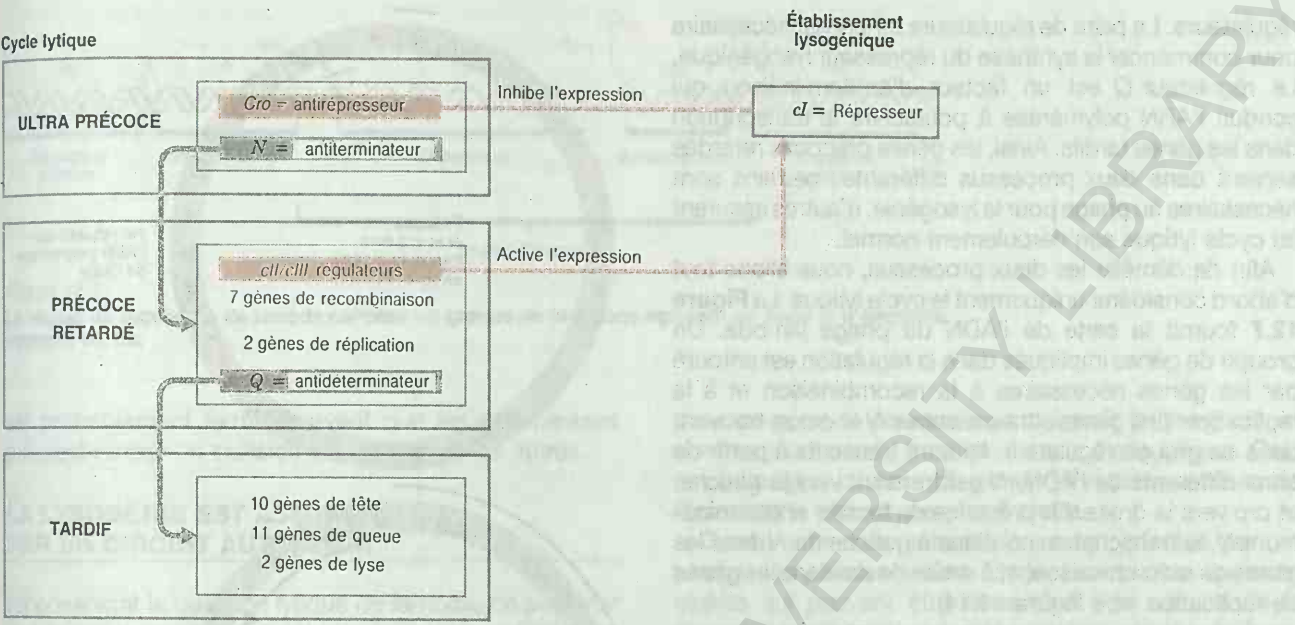


Figure 12.6 La cascade lytique de lambda est imbriquée avec le circuit d'établissement de la lysogénie.

Mais, comme cela est résumé dans la **Figure 12.6**, le déroulement du cycle lytique est imbriqué avec le circuit permettant de mettre en place la lysogénie.

Quand l'ADN du phage lambda rentre dans une nouvelle cellule hôte, les cycles lytique et lysogénique démarrent en même temps. Tous deux ont besoin de l'expression des gènes précoces immédiats et précoces retardés. Mais ensuite ils divergent, le développement lytique continue si les gènes tardifs sont exprimés, et la lysogénie se poursuit si la synthèse du répresseur a lieu.

Le phage lambda a seulement deux gènes ultra-précoces, transcrits indépendamment par l'ARN polymérase de l'hôte. Un de ceux-ci, le gène *cro*, a une double fonction : il empêche la synthèse du répresseur lysogé-

que (c'est une action indispensable si le cycle lytique est poursuivi) ; et il arrête l'expression des gènes ultra-précoces (qui ne sont plus nécessaires ultérieurement dans le cycle lytique). L'autre gène ultra-précoce est le gène *N*, qui, comme nous l'avons vu dans le Chapitre 11, code pour un facteur d'antitermination dont l'action au niveau des sites *nut* permet de poursuivre la transcription des gènes précoces retardés.

Les gènes précoces retardés contiennent deux gènes de réplication (nécessaires pour l'infection lytique), sept gènes de recombinaison (certains impliqués dans la recombinaison durant l'infection lytique, deux nécessaires, en cas de lysogénie, à l'intégration de l'ADN du phage lambda dans le chromosome bactérien), et trois gènes

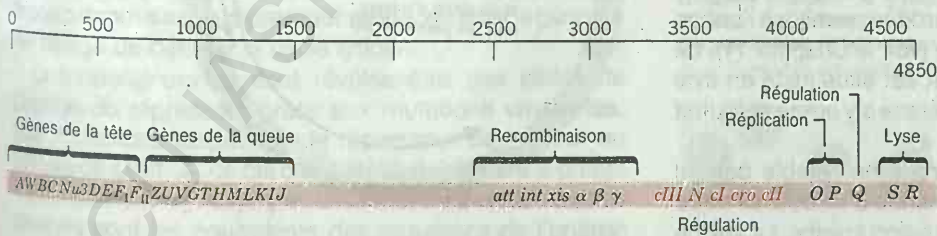


Figure 12.7 La carte de lambda montre des groupements de fonctions apparentées. Le génome fait 48 514 pb.

régulateurs. La paire de régulateurs *cII-cIII* est nécessaire pour commencer la synthèse du répresseur lysogénique. Le régulateur *Q* est un facteur d'antiterminalisation qui conduit l'ARN polymérase à poursuivre la transcription dans les gènes tardifs. Ainsi, les gènes précoces retardés servent dans deux processus différents : certains sont nécessaires au phage pour la lysogénie, d'autres assurent au cycle lytique son déroulement normal.

Afin de démêler les deux processus, nous allons tout d'abord considérer uniquement le cycle lytique. La **Figure 12.7** fournit la carte de l'ADN du phage lambda. Un groupe de gènes impliqués dans la régulation est entouré par les gènes nécessaires à la recombinaison et à la réplication. Les gènes ultra-précoces *N* et *cro* se trouvent dans ce groupe régulateur. Ils sont transcrits à partir de brins différents de l'ADN, *N* est transcrit vers la gauche, et *cro* vers la droite. En présence du facteur antiterminalisateur pN, la transcription continue à gauche de *N* dans les gènes de recombinaison, et à droite de *cro* dans les gènes de réplication (voir Figure 11.14).

La carte présentée dans la Figure 12.7 donne l'organisation de l'ADN du phage lambda tel qu'il existe dans la particule phagique. Mais précocement après l'infection, les extrémités de l'ADN se réunissent pour former un cercle. La **Figure 12.8** montre la véritable forme de l'ADN de lambda durant l'infection. Les gènes tardifs sont soudés en un seul groupe, contenant les gènes de lyse *S-R* à l'extrémité droite de la molécule d'ADN linéaire, et les gènes de la tête et de la queue *A-J* à l'extrémité gauche.

Les gènes tardifs sont exprimés comme une seule unité de transcription, commençant au promoteur P_R qui se trouve entre *Q* et *S*. Le promoteur tardif est utilisé constitutivement. Cependant, en l'absence du produit du gène *Q* (qui est le dernier gène vers la droite dans l'unité précoce retardée), la transcription tardive s'arrête au site t_{R3} . Le transcrit a 194 bases de long et il est appelé ARN 6S. Toutefois, quand pQ devient disponible, il supprime la terminaison au site t_{R3} ; l'ARN 6S est ainsi allongé conduisant les gènes tardifs à être largement exprimés. L'événement d'antiterminalisation tardif ressemble à l'action de pN à l'étape précoce retardée (voir le Chapitre 11). Le site *qut* au niveau duquel pQ agit est situé juste en aval du promoteur tardif, l'ARN polymérase s'y arrête et fournit une cible à l'action de pQ.

La transcription des gènes tardifs ne semble pas se terminer à un point précis, elle continue d'un bout à l'autre des gènes tardifs jusque dans la région voisine. Le même événement se produit avec la transcription précoce retardée vers la gauche qui continue après les fonctions de recombinaison. La transcription dans chaque direction

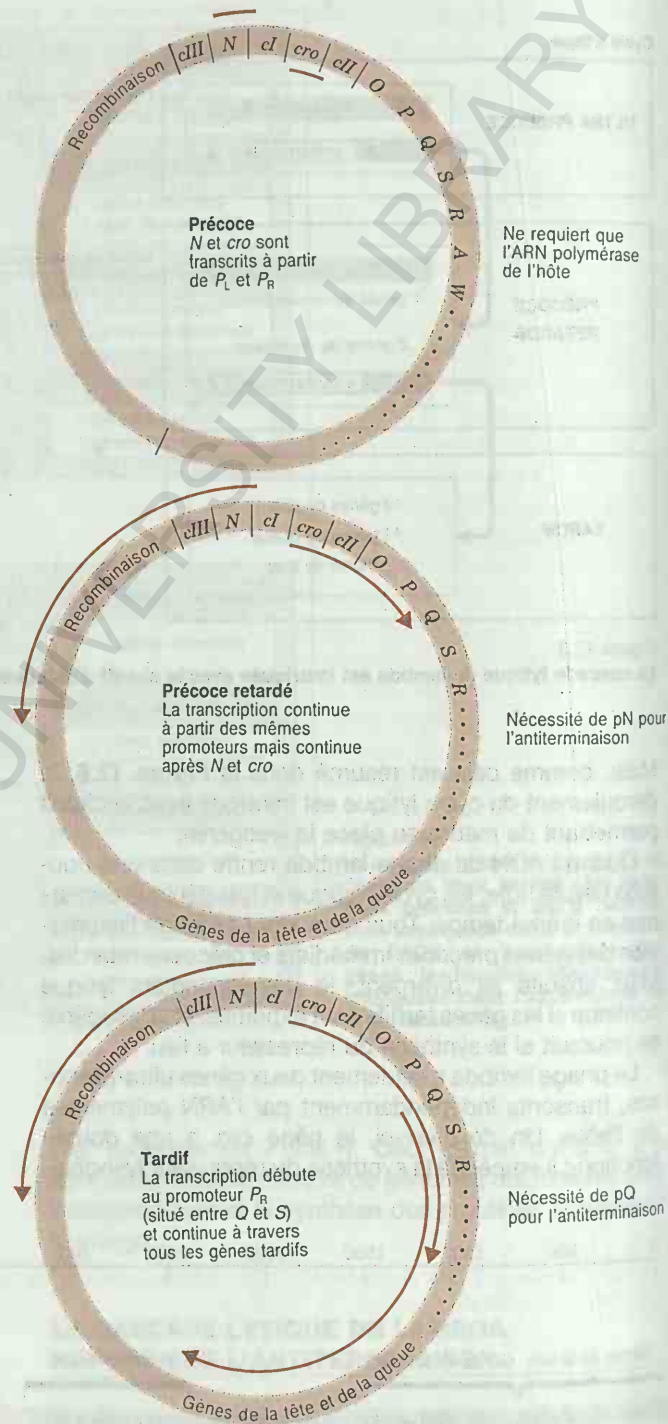


Figure 12.8

L'ADN du phage lambda se circularise durant l'infection ; ainsi, le groupe des gènes tardifs est intact dans une unité de transcription.

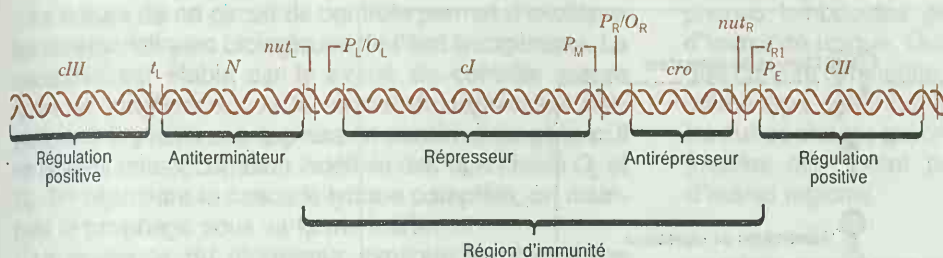


Figure 12.9

La région de régulation de lambda contient un groupe de fonctions agissant en *trans* et d'éléments agissant en *cis*.

est probablement terminée avant que les polymérases puissent rentrer en collision les unes avec les autres.

LA LYSOGÉNIE EST ASSURÉE PAR UN CIRCUIT AUTONOME

En observant la cascade lytique de lambda, on peut voir que le programme entier est mis en route par l'initiation de la transcription des gènes ultra-précoces *N* et *cro* au niveau des deux promoteurs P_L et P_R . Comme le phage lambda a recours à l'antitermination pour passer à l'étape d'expression suivante (précoce retardée), les deux mêmes promoteurs continuent à être utilisés tout au long de la phase précoce.

La carte détaillée de la région régulatrice présentée dans la Figure 12.9 montre que les promoteurs P_L et P_R se trouvent de part et d'autre du gène *cI*. A chacun des promoteurs est associé un opérateur (O_L , O_R) au niveau duquel une protéine répresseur se fixe, empêchant l'ARN polymérase d'initier la transcription. La séquence de chaque opérateur est chevauchante avec la séquence du promoteur qu'il contrôle. C'est pourquoi elles sont souvent appelées régions de contrôle P_L/O_L et P_R/O_R . En raison de la nature séquentielle de la cascade lytique, les régions de contrôle constituent une plaque tournante au niveau de laquelle l'entrée dans le cycle entier peut être contrôlée. En s'opposant à l'accès de l'ARN polymérase à ces promoteurs, le répresseur empêche ainsi le génome du phage de débiter le cycle lytique.

Les opérateurs se sont révélés être des cibles de l'action du répresseur grâce aux mutations **virulentes**. Ces mutations empêchent le répresseur de se fixer au niveau de O_L et O_R , ce qui conduit inévitablement le phage à poursuivre le cycle lytique. Les mutations virulentes des phages sont les équivalents des mutations de l'opéron entraînant une expression constitutive d'opérons bactériens.

Le répresseur est codé par le gène *cI*. Les mutants

dans ce gène ne peuvent pas maintenir la lysogénie, et sont toujours voués à entrer dans le cycle lytique. Ceci est à l'origine du nom du gène, qui reflète au cours de l'infection le phénotype **plage claire**.

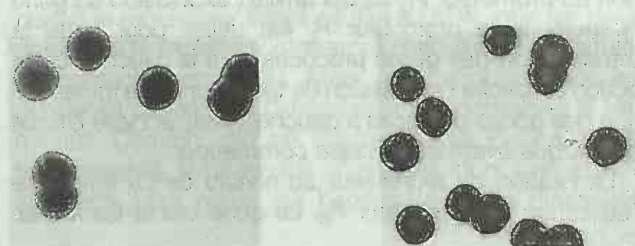
Quand une culture bactérienne est infectée par un phage, les cellules lysées forment de petites surfaces claires qui peuvent être observées à la surface de la plaque de culture et qui sont appelées **plages de lyse**. Avec des phages de type sauvage, les plages sont troubles car certaines cellules, au lieu d'être lysées, sont devenues lysogéniques. L'effet de la mutation *cI*⁻ est d'empêcher la lysogénie de telle sorte que les plages ne contiennent que des cellules lysées et apparaissent ainsi claires. La Figure 12.10 compare les deux types de plages.

Le gène *cI* est transcrit à partir du promoteur P_{RM} qui se trouve à son extrémité droite (l'indice « RM » signifie « repressor maintenance »). La transcription s'arrête à l'extrémité gauche du gène. Le messager commence en réalité avec le codon initiateur AUG ; du fait de l'absence de l'habituel site de fixation des ribosomes, l'ARNm est traduit de façon peu efficace, ce qui conduit à la production d'un faible niveau seulement de répresseur.

Figure 12.10

Les cultures de lambda de type sauvage forment des plages troubles (modèle de gauche) ; les mutants, qui ne peuvent pas devenir lysogènes, peuvent être détectés car ils produisent des plages claires (modèle de droite).

Avec l'aimable autorisation de Dale Kaiser.



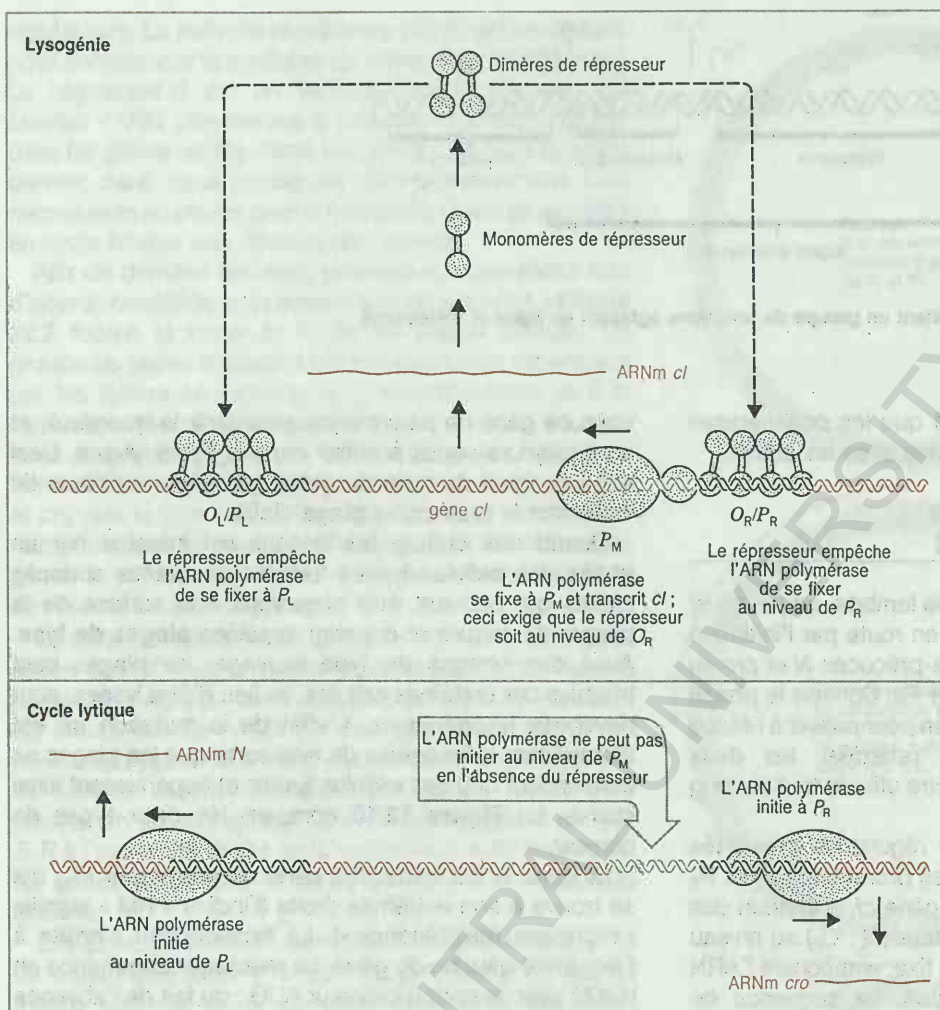


Figure 12.11

La lysogénie est maintenue par un circuit autogène (en haut). Si ce circuit est interrompu, le cycle lytique commence (en bas).

Le répresseur se fixe indépendamment au niveau des deux promoteurs. Il n'a qu'une seule fonction au niveau de O_L , mais une double fonction au niveau de O_R . Tout ceci est illustré dans la **Figure 12.11**.

Au niveau de O_L , le répresseur a le même type d'effet que ce qui a été décrit pour de nombreux autres systèmes : il empêche l'ARN polymérase d'initier la transcription au promoteur P_L , ce qui arrête l'expression du gène N , puisque le promoteur P_L est utilisé pour toute la transcription des gènes précoces vers la gauche ; cette action empêche l'expression de toute l'unité de transcription des gènes précoces à gauche. Ainsi, le cycle lytique est bloqué avant qu'il puisse commencer.

La fixation du répresseur au niveau de O_R empêche l'utilisation du promoteur P_R . Le gène cro et les autres

gènes précoces à droite ne peuvent donc pas être exprimés. (Nous verrons plus tard pourquoi il est important d'empêcher l'expression du gène cro quand la lysogénie doit être maintenue).

La présence du répresseur au niveau de O_R a aussi un autre effet. Le promoteur P_{RM} utilisé pour la synthèse du répresseur est adjacent à l'opérateur O_R . Il semble que l'ARN polymérase ne puisse initier à P_{RM} que lorsque le répresseur est lié à l'opérateur O_R . Le répresseur se comporte comme une protéine régulatrice positive nécessaire à la transcription du gène cl . Puisque le répresseur est le produit du gène cl , ces interactions créent un circuit autonome dans lequel la présence du répresseur est indispensable pour assurer de façon continue sa propre synthèse.

La nature de ce circuit de contrôle permet d'expliquer les caractéristiques biologiques de l'état lysogénique. La lysogénie est stable car le circuit de contrôle assure qu'aussi longtemps que le niveau du répresseur sera suffisant, il y aura une expression continue du gène *cI*. Il en résulte une occupation indéfinie des opérateurs O_L et O_R . En réprimant la cascade lytique complète, on maintient le prophage sous sa forme inerte.

La présence du répresseur explique le phénomène d'immunité. Si l'ADN d'un deuxième phage lambda rentre dans une cellule lysogénique, le répresseur qui est synthétisé par le génome du prophage résident, va immédiatement se fixer au niveau des opérateurs O_L et O_R du nouveau génome. Ceci empêche le deuxième phage de développer un cycle lytique.

Le prophage est induit à entamer le cycle lytique quand le circuit lysogénique est rompu. Ceci se produit quand le répresseur est inactivé. L'absence de répresseur permet à l'ARN polymérase de se fixer au niveau des promoteurs P_L et P_R , et d'initier le cycle lytique comme c'est illustré dans la Figure 12.11. La nature autonome du circuit de maintenance du répresseur conduit à une réponse très sensible. Comme la présence du répresseur est nécessaire à sa propre synthèse, l'expression du gène *cI* cesse dès que le répresseur existant est détruit. Il n'y a donc plus de synthèse de répresseur pour remplacer les molécules qui ont été endommagées. Ainsi, le cycle lytique peut-il commencer sans aucune interférence induite du circuit qui maintient la lysogénie.

La région comprenant les opérateurs droit et gauche, le gène *cI*, et le gène *cro* détermine l'immunité du phage. Tout phage qui possède cette région a le même type d'immunité, car cette région *spécifie à la fois le répresseur et le site au niveau duquel il agit*. Par conséquent, elle est appelée la **région d'immunité**. Chacun des quatre

phages lambda $\phi 80$, 21, 434, et λ a une région d'immunité unique. Quand on dit qu'un phage lysogénique confère l'immunité contre tous les autres phages du même type, cela veut dire, plus précisément, contre tous les autres phages qui ont une région d'immunité identique (quelles que soient par ailleurs les différences dans d'autres régions).

LE RÉPRESSEUR EST UN DIMÈRE QUI SE FIXE DE FAÇON COOPÉRATIVE AU NIVEAU DE CHAQUE OPÉRATEUR

La sous-unité du répresseur est un polypeptide de 27 000 daltons composé de deux domaines distincts comme cela est résumé dans la Figure 12.12.

- Le domaine N-terminal, qui va des résidus 1 à 92, contient le site de fixation pour l'opérateur.
- Le domaine C-terminal, qui va des résidus 132 à 236, est responsable de la formation du dimère.

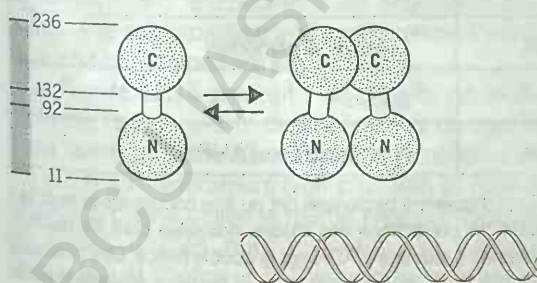
Les deux domaines sont reliés par une séquence de 40 résidus. Quand le répresseur est digéré par une protéase, chaque domaine est libéré en un fragment séparé.

Chaque domaine peut exercer sa fonction indépendamment de l'autre. Le fragment protéolytique C-terminal peut former des oligomères. Le fragment N-terminal peut fixer l'opérateur, mais avec une affinité plus faible que le répresseur intact. Donc, l'information de la spécificité de contact avec l'ADN est contenue dans le domaine N-terminal, mais (comme dans le cas du répresseur *lac*) l'efficacité est augmentée quand il est rattaché au domaine C-terminal.

La structure dimérique du répresseur est essentielle au maintien de la lysogénie. L'induction du prophage lysogénique à entrer dans un cycle lytique est produite par le clivage entre les résidus 110 et 111 de la région de connexion de la sous-unité du répresseur. (En un sens, on peut voir une certaine analogie avec le changement allostérique de conformation qui se produit quand la petite molécule inductrice inactive le répresseur d'un opéron bactérien, une propriété que ne possède pas le répresseur lysogénique.)

La dimérisation des domaines C-terminaux garantit, quand le répresseur intact se fixe à l'ADN, que les deux domaines N-terminaux rentreront chacun simultanément en contact avec l'ADN. Mais une coupure libère les domaines dimériques C-terminaux des domaines N-terminaux. Comme cela est illustré dans la Figure 12.13, les domaines N-terminaux n'ont plus alors une affinité suffisante pour rester sur l'opérateur. Ils se dissocient

Figure 12.12. Les régions N-terminales et C-terminales du répresseur forment des domaines séparés. Les domaines C-terminaux s'associent pour former un dimère ; les domaines N-terminaux fixent l'ADN.



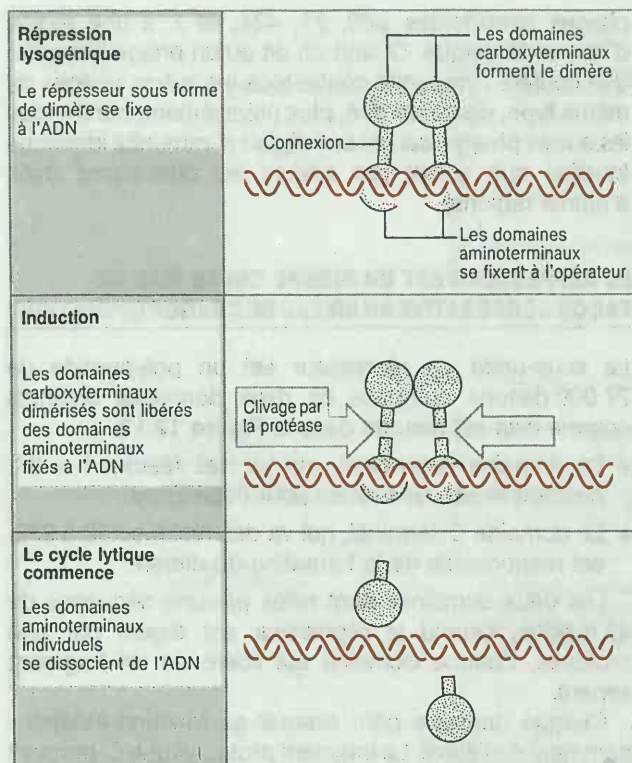


Figure 12.13

Le dimère de répresseur se fixe à l'opérateur. L'affinité des domaines N-terminaux pour l'ADN est contrôlée par la dimérisation des domaines C-terminaux.

donc de l'ADN, permettant ainsi à l'infection lytique de démarrer.

L'équilibre entre les cycles lysogénique et lytique dépend de la concentration du répresseur. Le répresseur intact est présent dans la cellule à une concentration

Tableau 12.1

Les domaines se fixant à l'ADN peuvent occuper des parties différentes de la protéine régulatrice.

Protéine	Taille du monomère	Forme active	Domaine liant l'ADN	Autre domaine
Répresseur λ	236	dimère	92 résidus N-terminaux	C-terminal (dimérise)
Cro	66	dimère	toute la protéine	aucun
Activateur CAP	209	dimère	74 résidus C-terminaux	N-terminal (fixe l'AMPc)

* N. d. T. : Depuis 1987, plusieurs répresseurs se liant à des opérateurs du phage λ ou d'opérons bactériens ont pu être cocrystallisés avec leur site de fixation sur l'ADN. Les résultats de l'étude cristallographique ont confirmé toutes les prédictions des études indirectes dont les résultats sont rapportés ici. Elles ont, de plus, permis de préciser l'influence de la liaison des protéines à leur site d'ADN sur la double hélice. Des courbures spontanées de l'ADN, probablement déterminées par des séquences nucléotidiques particulières (riches en A et en T) peuvent également influencer sur l'affinité des sites de liaison pour les protéines. En particulier, la nature des nucléotides séparant deux hémi-sites de fixation d'un opérateur intervient sur l'affinité pour le répresseur alors même qu'ils n'entrent pas en contact avec lui.

suffisante pour garantir que les opérateurs soient occupés. Mais si le répresseur est clivé, sa concentration devient insuffisante, en raison de la faible affinité des domaines N-terminaux séparés pour l'opérateur. Une trop grande concentration de répresseur pourrait rendre impossible l'induction du cycle lytique, mais un trop faible niveau de répresseur, évidemment, ne permettrait pas de maintenir la lysogénie.

Plusieurs protéines se liant à l'ADN, présentent un mode similaire de fixation à la double hélice *. Les caractéristiques générales de trois protéines de ce type sont résumées dans le **Tableau 12.1**. Bien que chaque protéine fonctionne sous forme dimérique, elles diffèrent par ailleurs dans leur organisation générale, puisque le répresseur de λ a son domaine N-terminal qui se lie à l'ADN, alors que c'est au niveau de son domaine C-terminal que la protéine CAP se lie à l'ADN et que la protéine Cro n'a qu'un seul domaine. Toutefois, dans chaque cas, le domaine actif contient de courtes régions en hélice α constituant le centre protéique en contact avec l'ADN.

Par exemple, le domaine N-terminal du répresseur λ contient plusieurs hélices α disposées comme cela est illustré dans la **Figure 12.14**. La structure de la connexion et du domaine C-terminal n'est pas connue.

Deux régions en hélice α sont responsables de la fixation à l'ADN, et le **modèle de contact des deux hélices** est illustré dans la **Figure 12.15**. Dans chaque monomère, la région en hélice α -3 est constituée de 9 acides aminés, formant un angle avec la précédente région de 7 acides aminés qui constitue l'hélice α -2. Dans le dimère, les deux régions en hélice α -3 se trouvent séparées de 34 Å ; elles peuvent s'emboîter dans deux grands sillons successifs de l'ADN. Les régions d'hélice α -2 forment un angle qui les place en travers de ce grand sillon.

Les séquences de ces régions sont partiellement

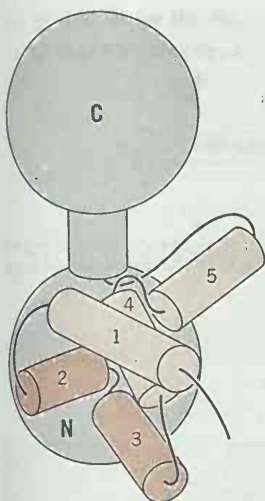


Figure 12.14

Le domaine N-terminal du répresseur de lambda contient cinq portions de l'hélice α . Les hélices α -2 et 3 font un angle entre elles et sont engagées dans la fixation avec l'ADN.

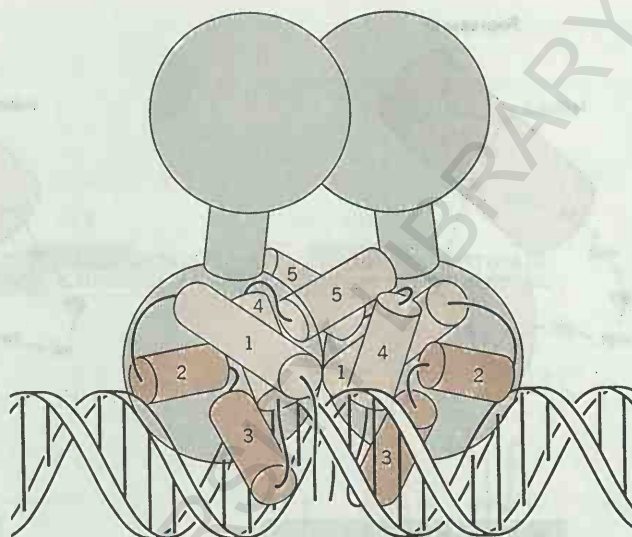


Figure 12.15

Dans le modèle à deux hélices, dans chaque monomère l'hélice α -3 se place dans le grand sillon de l'ADN et l'hélice α -2 se situe en travers du sillon.

homologues dans plusieurs protéines se fixant à l'ADN, dont la protéine CAP, le répresseur *lac*, et plusieurs répresseurs de phages. Elles sont responsables de la fixation aux opérateurs.

Les contacts entre l'hélice α -3 et l'ADN reposent sur des liaisons hydrogène entre les acides aminés de la chaîne polypeptidique et les paires de bases exposées. Cette région de la protéine est responsable de la reconnaissance de la séquence spécifique cible de l'ADN ; elle est donc appelée la **région de reconnaissance de l'hélice**.

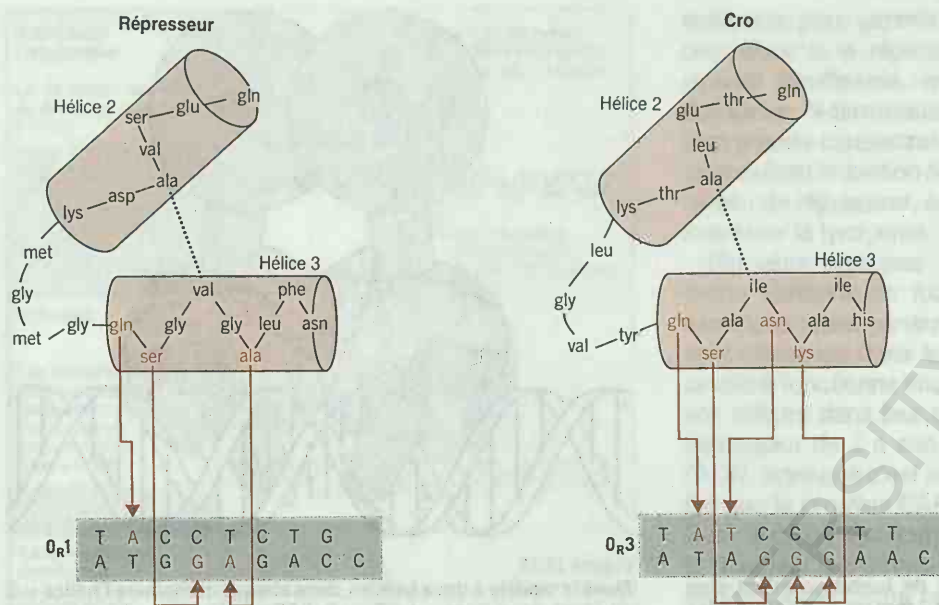
Si la région de reconnaissance de l'hélice d'un répresseur est remplacée par la séquence correspondante d'un autre répresseur (par manipulation de séquences codantes), la spécificité de la protéine hybride est celle de la nouvelle région de reconnaissance de l'hélice. *La séquence des acides aminés de cette courte région détermine donc la spécificité de séquence de chaque protéine, et est capable d'agir conjointement avec le reste de la chaîne polypeptidique.*

La **Figure 12.16** présente les détails de la fixation à l'ADN de l'hélice α -2 et de l'hélice α -3 du répresseur et d'une autre protéine de lambda, la protéine Cro qui se fixe aussi aux opérateurs (voir plus loin la Figure 12.24). Chaque protéine utilise le même type d'interaction entre les acides aminés pour maintenir les relations entre l'hélice α -2 et l'hélice α -3. Des acides aminés de l'hélice α -3 du répresseur établissent des contacts avec des bases spécifiques de l'opérateur. Dans l'hélice α -3 de la

protéine Cro, certaines des interactions sont les mêmes (ce qui explique pourquoi, le répresseur et la protéine Cro reconnaissent les opérateurs) et d'autres sont différentes (ce qui explique les différences d'affinité du répresseur et de la protéine Cro pour chaque opérateur). Nous voyons donc que les changements dans la séquence des acides aminés de l'hélice α -3 fournissent un réglage fin, contrôlant la spécificité de fixation à l'ADN avec une grande précision.

Les contacts de l'hélice α -2 avec l'ADN reposent sur des liaisons hydrogènes avec les phosphates du squelette. Ces interactions sont nécessaires pour la fixation, mais ne contrôlent pas la spécificité de reconnaissance. En plus de ces contacts, une grande partie de l'énergie générale d'interaction avec l'ADN provient des interactions ioniques avec les phosphates du squelette, qui ne dépendent pas de la séquence de l'ADN.

Le domaine N-terminal isolé du répresseur forme le même type de liaison avec l'ADN que la protéine intacte. Cependant, si l'on enlève les 6 derniers acides aminés N-terminaux, on élimine certaines de ces liaisons. Ces observations fournissent les bases d'un modèle qui est présenté dans la **Figure 12.17**, dans lequel la majorité du domaine N-terminal est en contact avec l'une des faces de l'ADN tandis que les 6 derniers acides aminés N-terminaux forment un bras qui s'étend vers l'autre face. L'interaction entre le bras et l'ADN contribue largement à la fixation à l'ADN. L'affinité pour l'ADN du répresseur sans bras est réduite d'un facteur 1000.

**Figure 12.16**

Deux protéines qui utilisent l'organisation en deux hélices pour interagir sur l'ADN reconnaissent les opérateurs de lambda avec des affinités qui sont fonction de la séquence des acides aminés de l'hélice α -3. Le profil des interactions est montré pour un site opérateur auquel chaque protéine se fixe très fortement.

Chaque opérateur possède trois sites de fixation pour le répresseur. Comme on peut le voir dans la **Figure 12.18**, chaque site de fixation est composé d'une séquence de 17 pb présentant une symétrie partielle dont l'axe passe par la paire de bases centrale. Il n'y a pas deux des six sites de fixation qui soient identiques, mais tous se conforment à une séquence consensus. Les sites de fixation à l'intérieur de chaque opérateur sont séparés par des zones de 3 à 7 paires de bases riches en paires de bases A-T. Les sites au niveau de chaque opérateur sont numérotés ; ainsi O_R est composé d'une série de sites de fixations O_{R1} - O_{R2} - O_{R3} alors que O_L est composé de la série O_{L1} - O_{L2} - O_{L3} . Dans chaque cas, le site 1 est le site le plus proche du point d'initiation de la transcrip-

tion au niveau du promoteur alors que les sites 2 et 3 se trouvent plus loin en amont.

A chaque site de fixation, l'empreinte des points de contact du répresseur suggère qu'il se fixe de façon symétrique ; ainsi chaque domaine N-terminal du dimère rentre en contact avec un ensemble similaire de bases. De cette façon, chaque région N-terminale se lie à un demi-site de fixation. Le côté du site de fixation, au niveau duquel le répresseur est en contact, est indiqué par des zones colorées dans la Figure 12.18. La Figure 12.16 montre l'interaction du répresseur avec l'un des demi-sites O_{R1} . Dans la structure en double hélice de l'ADN, les points de contact se trouvent essentiellement le long du grand sillon de l'ADN.

Comment le répresseur décide-t-il de l'endroit où se fixer, face aux trois sites de fixation possible au niveau de chaque opérateur ? Le site 1 de chaque opérateur a une plus forte affinité pour le répresseur (approximativement d'un facteur 10) que les autres sites. Le répresseur se fixe donc toujours en premier au niveau de O_{L1} et O_{R1} .

Le répresseur de λ se fixe à chaque opérateur de façon coopérative. La présence d'un dimère au site 1 augmente fortement l'affinité avec laquelle le second dimère peut se fixer au site 2. Quand les deux sites 1 et 2 sont

Figure 12.17

Une vue de derrière montre que la plus grande partie du répresseur interagit avec une face de l'ADN, mais son bras N-terminal rentre en contact avec l'autre face.



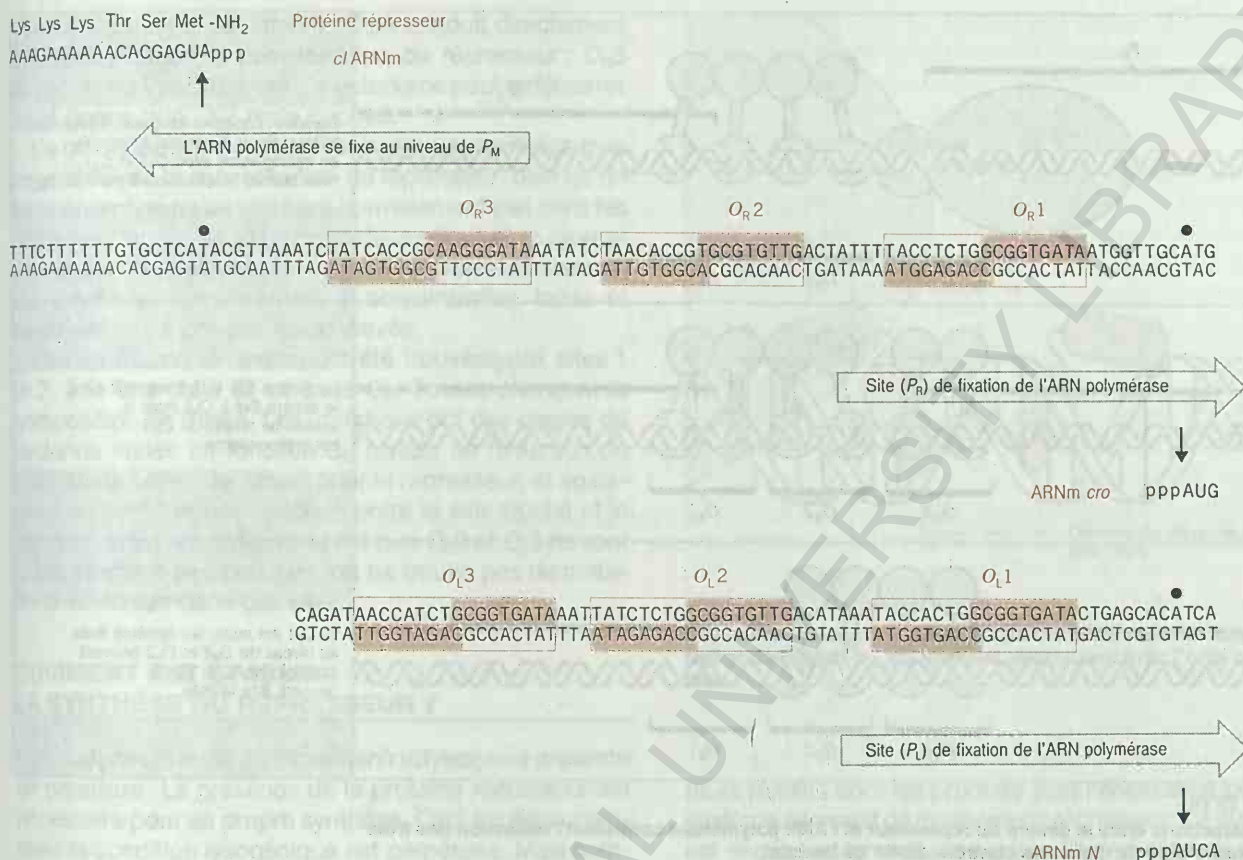


Figure 12.18

Chaque opérateur chevauche le promoteur auquel l'ARN polymérase se fixe et contient 3 sites de fixation pour le répresseur.

L'orientation de O_L a été inversée pour faciliter sa comparaison avec O_R.

occupés, cette interaction ne s'étend pas plus loin, vers le site 3. Aux concentrations de répresseur habituellement trouvées dans la cellule lysogénique, les deux sites 1 et 2 sont occupés au niveau de chaque opérateur, mais le site 3 reste inoccupé.

Si le site 1 est inactif (à cause d'une mutation), le répresseur se fixe alors coopérativement aux sites 2 et 3. Ainsi, la fixation au site 2 favorise donc la fixation d'un autre dimère au site 3. Cette interaction se produit directement avec le dimère du répresseur et non par l'intermédiaire d'un changement de conformation de l'ADN. Il est probable que la région de connexion oriente les régions C-terminales du premier répresseur de telle sorte qu'elles interagissent avec les régions C-terminales d'un second dimère.

A partir des séquences présentées dans la Figure 12.18, on peut voir que O_L1 et O_R1 se trouvent plus ou moins au centre des sites de fixation de l'ARN polymérase, respectivement sur les promoteurs P_L et P_R. L'occupation de O_L1-O_L2 et O_R1-O_R2 bloque donc physiquement l'accès à l'ARN polymérase aux promoteurs correspondants.

Une interaction différente existe entre O_R et le promoteur P_{RM} qui assure la transcription de *cl*. Le site de fixation de l'ARN polymérase est juste adjacent à O_R2. Ce qui explique comment le répresseur régule sa propre synthèse de façon autonome. Quand deux dimères sont fixés au niveau de O_R1-O_R2, le dimère au niveau de O_R2 interagit avec l'ARN polymérase, probablement par l'intermédiaire d'une interaction protéine-protéine. Contrai-

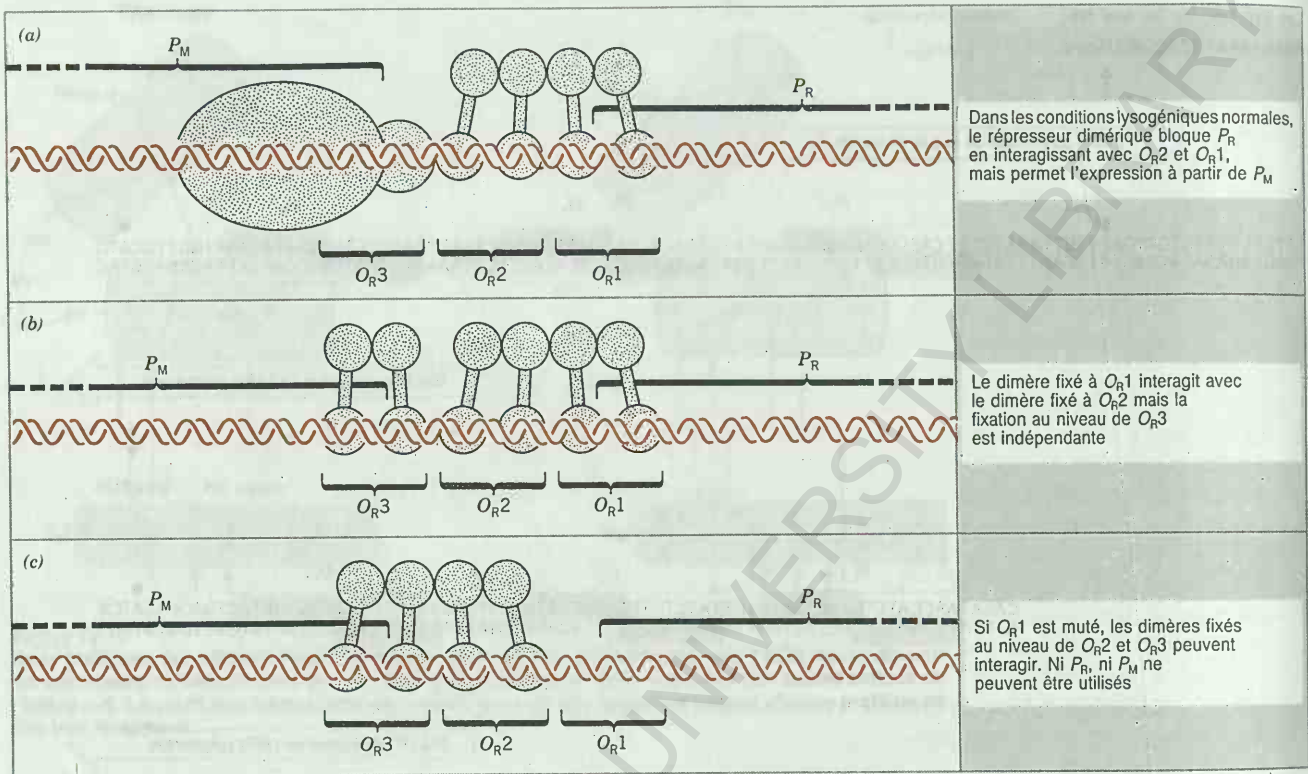


Figure 12.19

Les interactions entre le dimère du répresseur et l'ARN polymérase contrôlent l'utilisation des trois promoteurs dans la région de contrôle droite de lambda.

rement à l'interaction entre les répresseurs, cette propriété réside dans le domaine aminoterminal qui peut stimuler l'utilisation du promoteur P_{RM} , même sous forme d'un fragment indépendant. Cette interaction et d'autres au niveau de O_R/P_R sont illustrées dans la Figure 12.19. Le site O_R2 est seulement à une paire de bases plus près du point d'initiation de P_R que de celui de P_{RM} , ce qui entraîne néanmoins une différence fonctionnelle considérable.

Des mutations qui abolissent le contrôle positif sont localisées dans le gène *cl*. Ces mutants restent capables de se fixer à l'opérateur afin de réprimer la transcription, mais ne sont plus capables d'induire l'ARN polymérase à transcrire à partir du promoteur P_{RM} . Elles sont localisées dans un petit groupe d'acides aminés qui se situe soit dans l'hélice $\alpha-2$, soit dans l'angle que forment l'hélice $\alpha-2$ et l'hélice $\alpha-3$.

La localisation des mutations du contrôle positif du répresseur est indiquée dans la Figure 12.20. Le site est proche d'un groupe phosphate de l'ADN qui est à la fois reconnu par le répresseur et par l'ARN polymérase. Le

groupe d'acides aminés est donc dans une position adéquate dans le répresseur pour interagir avec l'ARN polymérase. L'interaction entre le répresseur et la polymérase est nécessaire pour que la polymérase effectue le passage du complexe fermé au complexe ouvert (voir le Tableau 12.2).

Les interactions protéine-protéine impliquées dans le contrôle positif peuvent revêtir différentes formes ; par exemple, des mutations analogues dans le phage P22 se trouvent à l'une des extrémités de l'hélice $\alpha-3$, alors que la structure se fixant à l'ADN du répresseur P22 est très similaire à celle du répresseur de λ . Un principe commun important est que l'énergie libérée lors de l'interaction protéine-protéine peut être utilisée pour réaliser des réactions nécessaires à l'initiation de la transcription.

Que se produit-il si le dimère du répresseur se fixe au niveau de O_R3 ? Ce site est chevauchant avec le site de fixation de l'ARN polymérase au niveau du promoteur P_{RM} . Si la concentration du répresseur devient suffisamment élevée pour entraîner l'occupation du site O_R3 , la

transcription de *cI* est inhibée. Cela conduit directement à la réduction de la concentration du répresseur ; O_R3 devient alors libre, et la boucle autonome peut redémarrer étant donné que O_R2 reste occupé.

Ce mécanisme pourrait prévenir une augmentation trop importante de la concentration du répresseur, bien qu'un tel phénomène ne se produise normalement pas dans les bactéries lysogènes. D'un point de vue formel, le répresseur est un régulateur autonome de sa propre expression qui fonctionne positivement à concentration faible et négativement à concentration élevée.

Des mutations virulentes ont été trouvées aux sites 1 et 2, à la fois dans O_L et dans O_R . Toutes changent la composition en bases. Les mutations ont des degrés de virulence variés en fonction du niveau de réduction de l'affinité des sites de liaison pour le répresseur, et également en fonction des relations entre le site touché et le promoteur. En accord avec le fait que O_R3 et O_L3 ne sont habituellement pas occupés, on ne trouve pas de mutations virulentes dans ces sites.

COMMENT EST ÉTABLIE LA SYNTHÈSE DU RÉPRESSEUR ?

Le circuit de contrôle pour maintenir la lysogénie présente un paradoxe. La présence de la protéine répresseur est nécessaire pour sa propre synthèse. Ceci explique comment la condition lysogénique est perpétuée. Mais comment la synthèse du répresseur est-elle établie au tout début ?

Quand un ADN de lambda rentre dans une nouvelle cellule hôte, l'ARN polymérase bactérienne ne peut transcrire *cI* parce que le répresseur n'est pas présent pour permettre la liaison à P_{RM} . Mais cette absence de répresseur veut également dire que P_R et P_L sont disponibles. Ainsi, le premier événement qui a lieu quand l'ADN de lambda infecte une bactérie, est la transcription des gènes *N* et *cro*. Ensuite, pN permet à la transcription d'être étendue au-delà. Ainsi, *cIII* (et d'autres gènes) sont transcrits sur la gauche, tandis que *cII* (et d'autres gènes) sont transcrits sur la droite (voir la Figure 12.9).

Les gènes *cII* et *cIII* partagent avec *cI* la propriété que des mutations en leur sein provoquent des plages claires. Mais il existe une différence. Les mutants *cI*⁻ ne peuvent ni établir, ni maintenir la lysogénie. Les mutants *cII*⁻ ou *cIII*⁻ ont de la difficulté à établir la lysogénie, mais une fois que celle-ci est établie, ils sont capables de la maintenir par le circuit autonome dont nous avons déjà discuté.

Ceci implique que les gènes *cII* et *cIII* sont des régula-

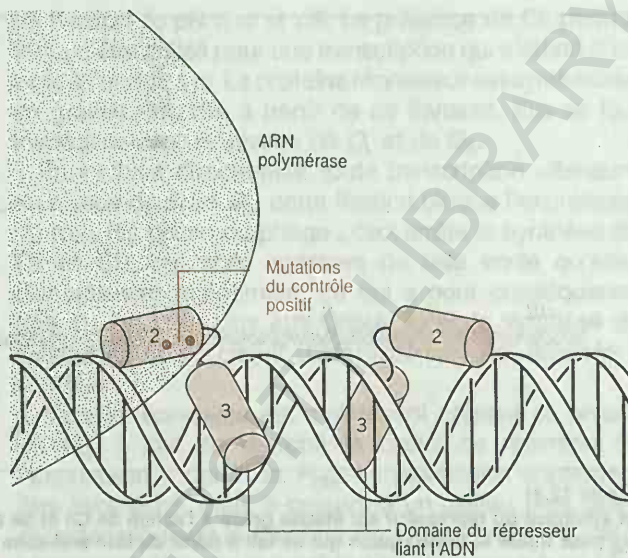


Figure 12.20

Des mutations du contrôle positif permettent d'identifier une petite région de l'hélice α -2 qui interagit directement avec l'ARN polymérase.

teurs positifs dont les produits sont nécessaires pour un système alternatif de synthèse de répresseur. Ce système est requis seulement pour initier l'expression de *cI* afin de pallier l'incapacité du circuit autonome à démarrer par une synthèse *de novo* de répresseur.

La protéine CII agit directement sur l'expression génique, tandis que CIII agit indirectement. Entre les gènes *cro* et *cII*, se trouve un autre promoteur appelé P_{RE} . (L'indice RE signifie « repressor establishment » établissement du répresseur.) Ce promoteur peut être reconnu par l'ARN polymérase seulement en présence de CII dont l'action est illustrée sur la Figure 12.21.

La protéine CII est extrêmement instable *in vivo*, car elle est dégradée sous l'effet de l'activité d'une protéine de l'hôte appelée HflA. Le rôle de CIII est de protéger CII contre la dégradation.

Le promoteur P_{RE} est dépourvu des deux séquences consensus normales. Cette déficience peut expliquer sa dépendance vis-à-vis de CII. Le développement d'un système de transcription *in vitro* au niveau de P_{RE} a rendu possible la définition du mode d'action de CII. Le promoteur est complètement inactif avec l'ARN polymérase seule, mais peut être transcrit lorsque CII est ajouté. Le régulateur se fixe à une région s'étendant de -25 à -45 environ. Quand l'ARN polymérase est ajoutée, une région

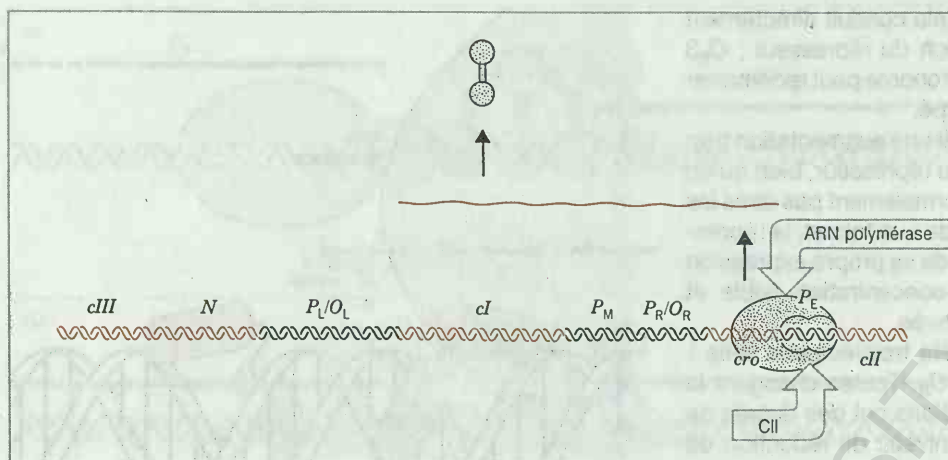


Figure 12.21

La synthèse du répresseur est établie grâce à l'action de CII et de l'ARN polymérase au niveau de P_{RE} pour initier la transcription qui se fait à partir du brin anti-sens de *cro* à travers le gène *cl*.

supplémentaire est protégée, s'étendant de -12 à $+11$. Comme le résume la Figure 12.22, les liaisons des deux protéines sont adjacentes ou chevauchantes.

L'importance des régions -35 et -10 pour la fonction promotrice, en dépit de leur absence de ressemblance avec la séquence consensus, est révélée par l'existence des mutations *cy*. Celles-ci ont des effets similaires à ceux de *cll*⁻ et *cll*⁺ en empêchant l'établissement de la lysogénie ; mais elles agissent en *cis* au lieu d'agir en *trans*. Elles sont réparties en deux groupes, *cyL* et *cyR* situés respectivement autour des positions -10 et -35 de P_{RE} .

Les mutations de *cyL* situées autour de -10 empêchent probablement l'ARN polymérase de reconnaître le promoteur. Les mutations *cyR* au niveau de -35 se répartissent en deux types. Dans la région correspondant à la boîte -35 , les mutations n'affectent pas la liaison de CII ; on suppose qu'elles empêchent la liaison de l'ARN polymérase. Bien que dépourvue d'homologie avec la séquence consensus, cette région est exactement comparable par sa taille et son emplacement.

De chaque côté de cette région, des mutations empêchent CII de se lier. Ces mutations touchent les courtes répétitions tétramériques TTGC. Chaque base du tétramère est séparée par 10 pb (un tour d'hélice) de son homologue dans l'autre tétramère de telle sorte que quand CII reconnaît les deux tétramères, il repose sur une des faces de la double hélice.

Le contrôle positif d'un promoteur implique qu'une protéine auxiliaire augmente l'efficacité avec laquelle l'ARN polymérase initie la transcription. Le tableau 12.2

Fixation de CII et de l'ARN polymérase en même temps

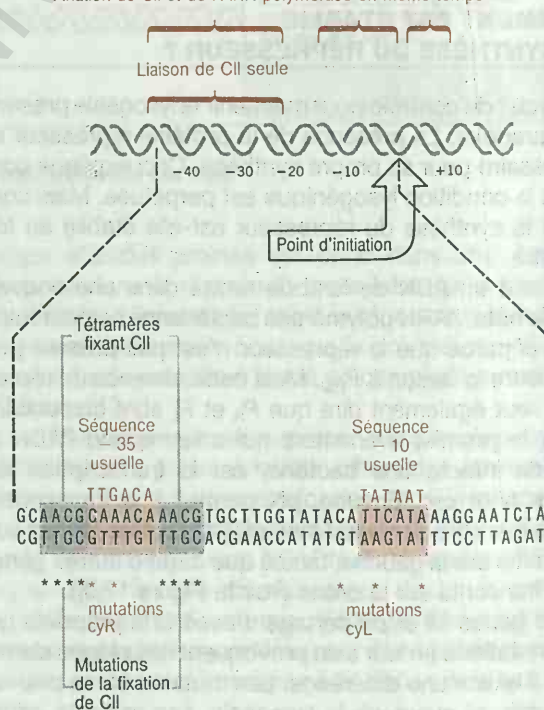


Figure 12.22

L'ARN polymérase se fixe au niveau du P_{RE} uniquement en présence de CII, qui se lie à la région aux environs de -35 . On notera que l'orientation du gène est présentée dans la direction 5'-3' habituelle, contrairement à ce qui est montré dans la figure 12.21.

Tableau 12.2
La régulation positive peut influencer l'ARN polymérase à n'importe quelle étape d'initiation de la transcription.

Locus	Régulateur	Liaison de la polymérase	Conversion fermé-ouvert
<i>lac</i>	AMP cyclique	20x	pas d'effet
λP_{RM}	répresseur	pas d'effet	11x
λP_{RE}	CII	100x	100x
P22 P_{MNT}	?	pas d'effet	pas d'effet

La liaison de l'ARN polymérase est mesurée par la constante de vitesse (K_a) par laquelle l'enzyme se lie au promoteur pour former un complexe fermé. La conversion fermé-ouvert est mesurée par la constante de vitesse (K_o) pour la conversion de ce complexe en la forme ouverte active dans laquelle la fusion initiale de l'ADN a eu lieu. Le chiffre représente le facteur d'augmentation de la constante de vitesse. Dans le promoteur P22, la régulation positive n'implique aucun changement dans les vitesses de l'ARN polymérase, mais est due à un effet indirect qui expulse l'enzyme d'un site adjacent.

résume les données, en montrant que chacune des deux étapes de l'interaction promoteur-polymérase — la liaison initiale pour former un complexe fermé, ou sa conversion en un complexe ouvert — peut être la cible de la régulation.

Le transcrit P_{RE} contient le brin **anti-sens** du gène *cro*, qui est habituellement transcrit dans la direction opposée (c'est-à-dire à partir de P_R). La séquence anti-sens *cro* n'est pas traduite sur le transcrit P_{RE} , mais la région codante *cI* est traduite très efficacement (en contraste avec la faible traduction du transcrit de P_{RM} mentionnée précédemment). En fait, le répresseur est synthétisé de 7 à 8 fois plus efficacement en utilisant P_{RE} plutôt que P_{RM} .

Il semble curieux que le gène *cro* puisse être transcrit simultanément dans des directions opposées. Nous ne savons vraiment pas ce qui arrive si deux ARN polymérases se rencontrent lorsqu'elles se déplacent l'une vers l'autre. On présume raisonnablement qu'un enchevêtrement se produirait, bloquant la transcription, mais il est possible que les polymérases soient capables de passer d'une manière que nous ne comprenons pas. L'explication peut être que, comme le promoteur P_{RE} est plus fort, son utilisation empêche l'utilisation de P_R .

A présent, nous pouvons voir comment la lysogénie est établie lors d'une nouvelle infection. La **Figure 12.23** récapitule les stades initiaux de la transcription de *N* et *cro* et l'extension de cette transcription par l'intermédiaire

de l'action de pN à *cI* et *cIII*. La présence de CII permet à P_{RE} d'être utilisé pour une transcription qui s'étend d'un bout à l'autre de *cI*. La protéine répresseur est synthétisée en grande quantité à partir de ce transcrit. Elle se fixe immédiatement au niveau de O_L et de O_R .

En inhibant directement toute transcription ultérieure au niveau de P_L et P_R , cette fixation bloque l'expression de tous les gènes du phage ; ceci arrête la synthèse de CII et CIII, qui sont instables de telle sorte qu'elles disparaissent rapidement, ce qui a pour conséquence que P_{RE} ne peut plus être utilisé. Ainsi, la synthèse de répresseur lié au circuit d'établissement est amenée à s'arrêter.

Mais le répresseur est maintenant présent au niveau de O_R . Il met en marche le circuit de maintien de l'expression à partir de P_{RM} . Le répresseur continue à être synthétisé, mais à présent à un niveau plus faible caractéristique de la fonction P_{RM} . Ainsi, le circuit d'établissement démarre la synthèse de répresseur à un haut niveau ; puis le répresseur bloque toutes les autres fonctions tandis qu'il met en marche dans le même temps le circuit de maintenance fonctionnant au faible niveau qui est efficace pour maintenir la lysogénie.

Nous n'allons pas voir en détail maintenant les autres fonctions nécessaires pour établir la lysogénie, mais nous pouvons seulement faire remarquer en quelques mots que l'ADN de lambda qui infecte doit être inséré dans le génome bactérien (voir Chapitre 28). L'insertion requiert le produit du gène *int*, qui est exprimé à partir de son propre promoteur P_i , au niveau duquel CII est aussi nécessaire. La séquence de P_i révèle des homologies avec P_{RE} dans le site de liaison de CII (mais pas dans la région - 10). Les fonctions nécessaires pour l'établissement du circuit de contrôle lysogénique sont donc sous le même contrôle que les fonctions requises pour l'intégration de l'ADN. Ainsi, l'établissement de la lysogénie dépend d'un contrôle qui assure que tous les événements nécessaires ont lieu avec la même chronologie.

L'ANTIRÉPRESSEUR EST REQUIS POUR L'INFECTION LYTIQUE

Nous avons commencé ce chapitre en disant que lambda possède les alternatives de devenir lysogène ou de démarrer une infection lytique. Nous avons vu que la lysogénie était initiée en établissant un circuit de maintien autonome, qui inhibe toute la cascade lytique en agissant en deux endroits. Le programme d'établissement de la

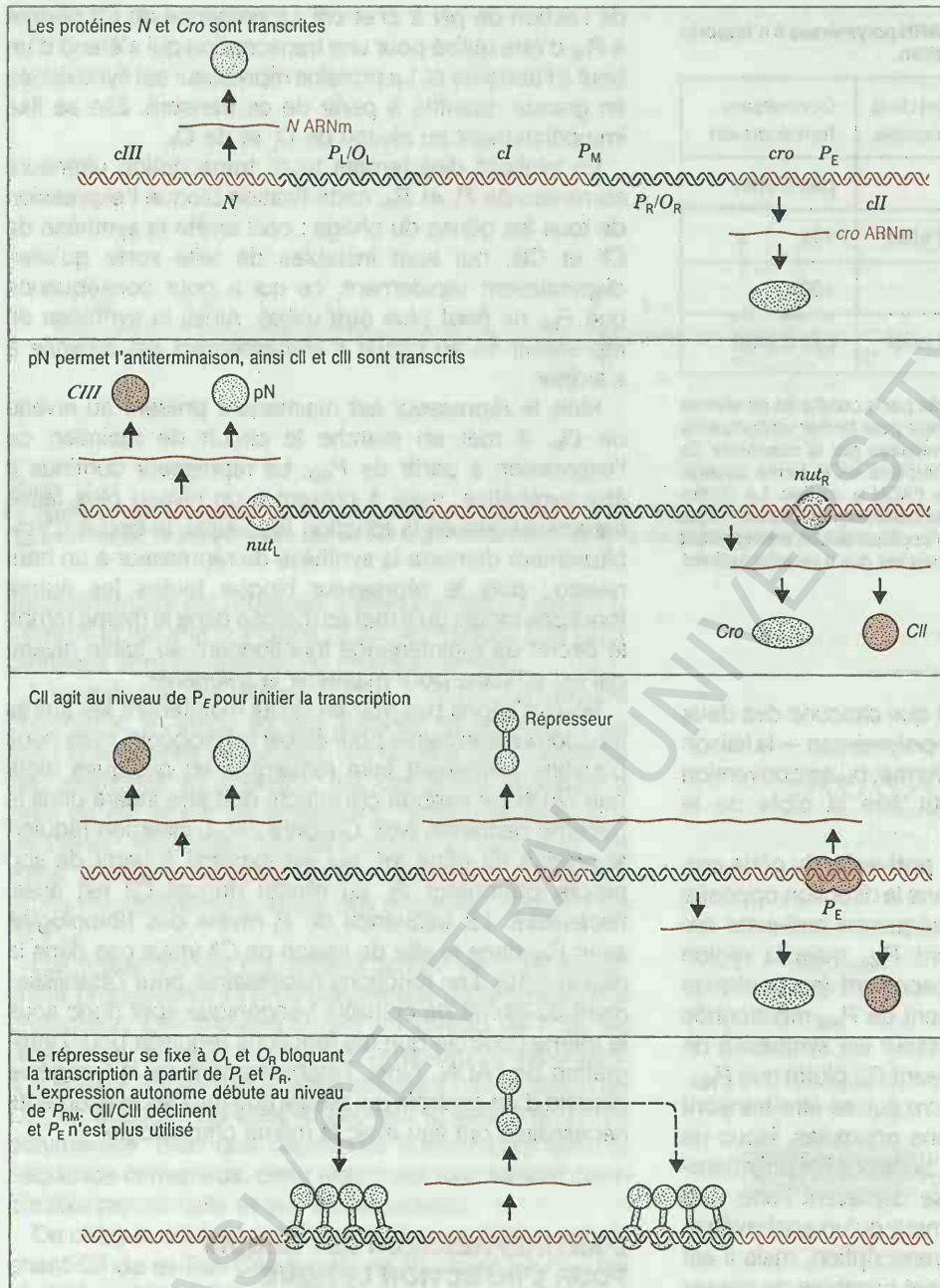


Figure 12.23

Une cascade est nécessaire pour établir la lysogénie, mais ensuite ce circuit est arrêté et remplacé par le circuit autonome de maintenance du répresseur.

lysogénie procède en fait par certains des événements que nous avons décrits précédemment en terme de cascade lytique. (L'expression de gènes précoces retardés via l'expression de *N* est requise.) Nous nous trouvons

maintenant face à un problème. Comment le phage entre-t-il dans le cycle lytique ?

Ce que nous avons laissé de côté jusqu'à présent est le rôle du gène *cro*, qui code pour une autre protéine

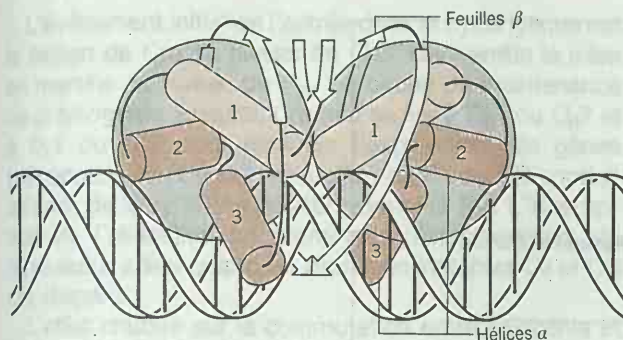


Figure 12.24

La protéine Cro est constituée d'un seul domaine qui contient des feuillets β et des hélices α ; un motif hélice-tour-hélice se lie à l'opérateur de la même façon que le fait le répresseur.

régulatrice parfois connue sous le nom d'**antirépresseur**. Ce régulateur est capable d'empêcher la synthèse de la protéine répresseur ; cette action supprime donc la possibilité d'établir la lysogénie. Des mutants de type *cro* établissent habituellement la lysogénie plutôt que d'entrer dans la voie lytique, car ils n'ont pas la capacité d'échapper au répresseur.

Cro forme un petit dimère (la sous-unité fait 9 000 daltons) qui agit à l'intérieur de la région d'immunité. Il a deux effets : il empêche la synthèse de répresseur par l'intermédiaire du circuit de maintenance, c'est-à-dire qu'il empêche la transcription par l'intermédiaire de P_{RM} . Il inhibe aussi l'expression des gènes précoces à partir de P_L et P_R . Ceci veut dire que quand un phage entre dans la voie lytique, Cro est responsable à la fois de l'inhibition de la synthèse de répresseur et de la diminution de l'expression des gènes précoces.

Cro accomplit sa fonction en se liant aux opérateurs. Il se fixe de manière symétrique à ses sites, réalisant des contacts similaires à ceux qui sont réalisés par le répresseur. En fait, Cro comprend une région qui possède la même structure générale que le répresseur ; une hélice α -2 fait un angle avec l'hélice α -3. L'interaction de la région hélice-tour-hélice avec l'ADN est illustrée dans la Figure 12.24.

Les séquences de Cro et du répresseur dans la région hélice-tour-hélice sont partiellement homologues, ce qui explique leurs capacités à rentrer en contact avec les mêmes séquences d'ADN (voir la Figure 12.16). Cependant, Cro rentre en contact avec une seule face de l'ADN ; il est dépourvu des bras N-terminaux qui permettent au répresseur d'atteindre l'autre face.

Comment deux protéines peuvent-elles avoir les mêmes sites d'action et cependant des effets si oppo-

sés ? La réponse tient dans les différences d'affinité qu'a chaque protéine pour les différents sites de liaison à l'intérieur des opérateurs. Considérons simplement O_R pour lequel on connaît le plus de choses et sur lequel Cro exerce ses deux effets. La série d'événements est illustrée dans la Figure 12.25.

L'affinité de Cro pour O_{R3} est plus grande que son affinité pour O_{R2} et O_{R1} . Ainsi, il se fixe d'abord à O_{R3} , ce qui empêche l'ARN polymérase de se fixer à P_{RM} . La première action de Cro est ainsi d'empêcher le circuit de maintenance de la lysogénie de jouer son rôle.

Puis Cro se fixe à O_{R2} ou O_{R1} . Son affinité pour ces sites est similaire et il n'y a pas d'effet coopératif. Sa présence à chacun des sites est suffisante pour empêcher l'ARN polymérase d'utiliser P_R . Ceci, à son tour, arrête la production des fonctions précoces (qui inclut la fonction Cro elle-même). Comme CII est instable, l'utilisation de P_{RE} est arrêtée. Ainsi, les deux actions de Cro bloquent ensemble toute production de répresseur.

En ce qui concerne le cycle lytique, Cro abaisse (bien qu'il ne l'élimine pas complètement) l'expression des gènes précoces. Son effet incomplet est expliqué par son affinité pour O_{R1} et O_{R2} qui est environ 8 fois plus faible que celle du répresseur. Cet effet de Cro n'a lieu que quand les gènes précoces sont devenus plus ou moins superflus, en raison de la présence de pQ ; à ce moment, le phage a débuté son expression génétique tardive et il s'applique à produire les particules de phages de sa descendance.

UN ÉQUILIBRE DÉLICAT : LA LYSOGÉNIE OU LA LYSE

Les programmes des voies lysogéniques et lytiques sont si intimement liés qu'il est impossible de prévoir le devenir d'un génome de phage individuel quand il entre dans une nouvelle bactérie hôte. L'antagonisme entre le répresseur et Cro se résoudra-t-il par l'établissement du circuit autonome de maintenance montré dans la Figure 12.23, ou bien par l'arrêt de la synthèse du répresseur et l'entrée dans la phase tardive de développement, comme l'illustre la Figure 12.25 ?

La même voie est suivie dans les deux cas jusqu'au moment de la décision. Les deux processus impliquent l'expression des gènes ultra-précoces et l'extension vers les gènes tardifs retardés. La différence entre ces deux alternatives survient jusqu'au moment où il s'agit de savoir si c'est Cro ou le répresseur qui va occuper les deux opérateurs.

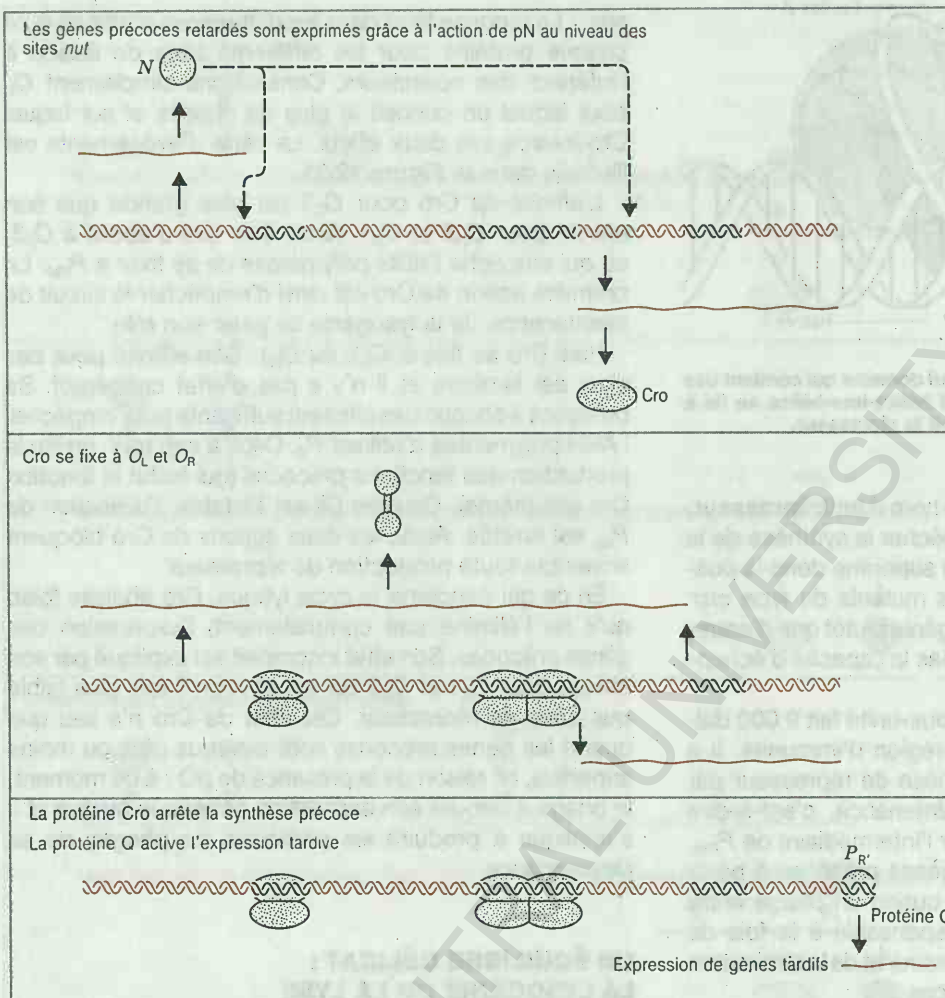


Figure 12.25

La cascade lytique nécessite la protéine Cro qui, directement, empêche l'expression du répresseur par l'intermédiaire de P_{RM} et, en inhibant l'expression des gènes précoces retardés, prévient indirectement l'établissement du répresseur par l'intermédiaire de P_{RE} .

La phase précoce au cours de laquelle la décision est prise est de durée limitée. Indépendamment de la voie que le phage va suivre, l'expression de tous les gènes précoces sera empêchée car P_L et P_R sont réprimés ; et la conséquence de la disparition de CII et CIII est l'arrêt de la production de répresseur via P_{RE} .

La question cruciale est de savoir si l'arrêt de transcription à partir de P_{RE} est suivi par l'activation de P_{RM} et l'établissement de la lysogénie ou si P_{RM} n'arrive pas à

devenir actif, le régulateur pQ engageant alors le phage dans la voie lytique.

L'événement initial dans l'établissement de la lysogénie est la liaison du répresseur au niveau d' O_{L1} et d' O_{R1} . La liaison au niveau de ces premiers sites est rapidement suivie par la liaison coopérative d'un autre dimère de répresseur au niveau d' O_{L2} et d' O_{R2} . Ceci arrête la synthèse de Cro et met en route la synthèse de répresseur via P_{RM} .

L'événement initial de l'entrée dans le cycle lytique est la liaison de Cro au niveau de O_{R3} . Ceci arrête la mise en marche au niveau de P_{RM} du circuit de maintenance de la lysogénie. Ensuite, Cro doit se lier à O_{R1} ou O_{R2} et à O_L1 ou O_L2 pour diminuer l'expression des gènes précoces. L'arrêt de la production de CII ou CIII conduit à l'arrêt de la synthèse de répresseur via P_{RE} . L'interruption de l'utilisation du promoteur d'établissement du répresseur a lieu quand les protéines instables CII et CIII ont disparu.

L'effet critique sur la commutation entre lysogénie et lyse est lié à CII. Si CII est active, la synthèse de répresseur via le promoteur d'établissement est efficace ; et en conséquence, le répresseur occupe l'opérateur. Si CII n'est pas active, l'établissement du répresseur échoue et Cro se fixe aux opérateurs.

Le niveau de protéine CII dans un ensemble particulier de circonstances est ce qui va déterminer le devenir de l'infection. Des mutations qui augmentent la stabilité de CII augmentent la fréquence de la lysogénie. De telles mutations peuvent avoir lieu dans *cII* elle-même ou dans d'autres gènes. La cause de l'instabilité de CII est sa susceptibilité à la dégradation par des protéases de l'hôte. Son niveau dans la cellule est influencé par une autre fonction de lambda, de même que par des fonctions de l'hôte.

L'effet de la protéine CIII de lambda est secondaire ; elle aide à protéger CII de la dégradation. Bien que la présence de CIII ne garantisse pas la survie de CII, en l'absence de CIII, CII est pratiquement toujours inactivée.

Des mutations dans les gènes de l'hôte *hflA* et *hflB* augmentent la lysogénie – *hfl* signifie haute fréquence de lysogénie. Les mutations stabilisent CII probablement parce qu'elles inactivent une ou des protéases de l'hôte dont CII est la cible. L'influence de la cellule hôte sur le niveau de CII peut fournir un moyen pour la bactérie d'interférer avec la prise de décision. Par exemple, des protéases de l'hôte qui dégradent CII sont activées par la croissance en milieu riche, de telle manière que lambda a tendance à lyser les cellules qui poussent bien, mais il a plus de chance de devenir lysogène dans des cellules carencées (et qui peuvent manquer des constituants nécessaires à une croissance lytique efficace).

POUR EN SAVOIR PLUS

Un examen initial des fonctions du répresseur de lambda a été réalisé par Ptashne (pp. 325-343) dans *The Operon* (Miller & Reznikoff, Eds., Cold Spring Harbor Laboratory, New York, 1978) ; il a depuis passé en revue tout ce qui a été fait concernant la régulation du répresseur et de la Cro dans *A Genetic Switch* (Cell Press & Blackwell Scientific Publications, San Francisco, 1986). Le cycle de lambda a été passé en revue par Friedman & Gottesman (pp. 43-51) dans *Lambda II* (Hendrix, Roberts, Stahl & Weisberg, Eds., Cold Spring Harbor Laboratory, New York, 1983). Les interactions entre lambda et son hôte ont été passées en revue par Friedman et al. (*Microbiol. Rev.* 48, 299-325, 1984).

PARTIE 4

LA PERPÉTUATION DE L'ADN

S'il est vrai que l'essence de la vie réside dans l'accumulation des expériences au cours des générations, l'on peut alors suspecter que le problème clé de la biologie, du point de vue d'un physicien, est de savoir comment la matière vivante enregistre et perpétue ces expériences. Prenons l'exemple d'une bactérie se trouvant dans un grand volume de milieu dont la composition chimique est favorable à sa croissance. Elle se nourrit, grossit puis se divise en deux. Les deux cellules filles vont elles aussi se diviser, comme le manche à balai de l'Apprenti Sorcier. Occasionnellement, une des répliques est légèrement imparfaite, il y a alors création d'un individu ayant des propriétés différentes, qui va se multiplier sous cette forme modifiée. Tout porte à croire que le jeu de l'évolution est le fruit de la reproduction qui, par le biais des mutations, peut donner lieu à une variété infinie d'individus.

Max Delbruck, 1949

PARTIE 4

LA PERPÉTUATION DE L'ADN

Il est évident que l'existence de la vie n'est pas une simple question de hasard. Elle est le résultat d'une série de processus complexes qui ont permis à la vie de se maintenir et de se développer. Ces processus sont régis par des lois physiques et chimiques qui sont les mêmes partout dans l'univers. C'est pourquoi la vie est si commune dans l'univers. Elle est partout où les conditions sont favorables. C'est pourquoi la vie est si ancienne. Elle a existé depuis des milliards d'années. C'est pourquoi la vie est si diversifiée. Elle a pris de nombreuses formes différentes. C'est pourquoi la vie est si fascinante. Elle est le résultat d'une série de processus complexes qui ont permis à la vie de se maintenir et de se développer. Ces processus sont régis par des lois physiques et chimiques qui sont les mêmes partout dans l'univers. C'est pourquoi la vie est si commune dans l'univers. Elle est partout où les conditions sont favorables. C'est pourquoi la vie est si ancienne. Elle a existé depuis des milliards d'années. C'est pourquoi la vie est si diversifiée. Elle a pris de nombreuses formes différentes. C'est pourquoi la vie est si fascinante.

Paris, 1955

CHAPITRE 13

LE RÉPLICON : UNITÉ DE RÉPLICATION

La totalité du génome doit être répliquée une fois, de façon précise, à chaque division cellulaire et ce, que la cellule soit procaryote (un seul chromosome) ou eucaryote (plusieurs chromosomes).

Deux relations corrélient réplication et division cellulaires :

- L'initiation de la réplication engage la cellule (procaryote ou eucaryote) dans une future division. C'est donc la décision d'initier ou non la réplication qui va déterminer le nombre de descendants d'une cellule.
- La réplication de l'ADN doit être terminée pour que la division cellulaire ait lieu. En effet, ce n'est qu'à la fin de la réplication que se produit la ségrégation du génome dupliqué dans chacune des cellules filles (par la voie de la mitose chez les eucaryotes). L'unité de ségrégation est le chromosome.

L'unité d'ADN où se produit la réplication est appelée le **réplicon**. Chaque réplicon n'est requis qu'une seule fois par cycle cellulaire. Le réplicon possède les éléments qui contrôlent la réplication : il a une **origine** où est initiée

la réplication, il peut aussi avoir un site de **terminaison** où est arrêtée la réplication.

Toute séquence qui n'est pas séparée d'une origine par un site de terminaison est répliquée comme faisant partie de ce réplicon. L'origine est un site en *cis* dont l'action est limitée à la molécule d'ADN à laquelle il est physiquement lié.

Le contrôle de la réplication s'effectue au niveau de l'initiation. *En effet, une fois initiée, la réplication continue jusqu'à duplication complète du réplicon.* La fréquence d'initiation est elle-même contrôlée par l'interaction de protéine(s) régulatrice(s) avec l'origine.

Au sens premier du terme, le réplicon (chez les procaryotes) était défini comme possédant à la fois l'origine de réplication et le gène codant pour la protéine régulatrice. Aujourd'hui cependant, le terme de réplicon est souvent utilisé chez les eucaryotes pour définir une unité de réplication contenant simplement une origine, toute(s) protéine(s) régulatrice(s) nécessaire(s) pouvant être codée(s) par un autre réplicon.

Le chromosome bactérien constitue à lui seul un réplicon. Dans ce cas, unité de ségrégation et unité de réplication ne font qu'une. L'initiation à une seule origine permet la réplication du génome entier.

En plus de leur chromosome, les bactéries peuvent abriter des plasmides ; un plasmide est un élément constitué d'ADN circulaire et dont la réplication est autonome. De même que les plasmides, tous les phages ou virus à ADN sont considérés comme des réplicons, ces derniers pouvant initier la réplication à plusieurs reprises pendant le cycle lytique.

Plus clairement, le réplicon procaryote peut être défini comme toute molécule d'ADN contenant une origine et pouvant se répliquer de façon autonome dans la cellule. La fréquence de réplication dépend de l'interaction des protéines régulatrices avec l'origine.

Un chromosome eucaryote contient un grand nombre de réplicons, ce qui signifie qu'une unité de ségrégation est constituée de plusieurs unités de réplication. Cette notion apporte une autre dimension au contrôle de la réplication. En effet, durant le cycle cellulaire, tous les réplicons vont être activés de manière asynchrone pendant des périodes équivalentes ; chacun d'entre eux n'est cependant activé qu'une seule fois par cycle cellulaire. *Il doit donc exister un signal permettant la distinction*

entre réplicons déjà répliqués et non répliqués afin d'éviter qu'ils ne s'activent une seconde fois ; par ailleurs, un autre signal doit indiquer que la totalité de la réplication a été effectuée.

Quelles sont les séquences qui font office d'origine de réplication et comment sont-elles reconnues par les protéines régulatrices de la réplication ? En premier lieu et d'un point de vue moléculaire, il s'agit donc de savoir si c'est une séquence précise d'ADN ou bien une structure secondaire que la molécule adopte par endroits, qui est reconnue comme origine de réplication. En second lieu, il s'agit de comprendre comment les protéines régulatrices initient un et un seul cycle de réplication par division cellulaire.

LA RÉPLICATION SÉQUENTIELLE FORME DES YEUX

Considérons une molécule d'ADN engagée dans la réplication. Comme l'illustre la **Figure 13.1**, la région qui n'est pas encore répliquée est constituée des brins parentaux qui vont être séparés afin que la synthèse des molécules filles se poursuive. L'endroit où se fait la réplication est appelé **fourche de réplication** (connu aussi sous le nom

Figure 13.1

La réplication peut être unidirectionnelle ou bidirectionnelle selon qu'il y a formation d'une ou deux fourches de réplication à partir de l'origine du réplicon.

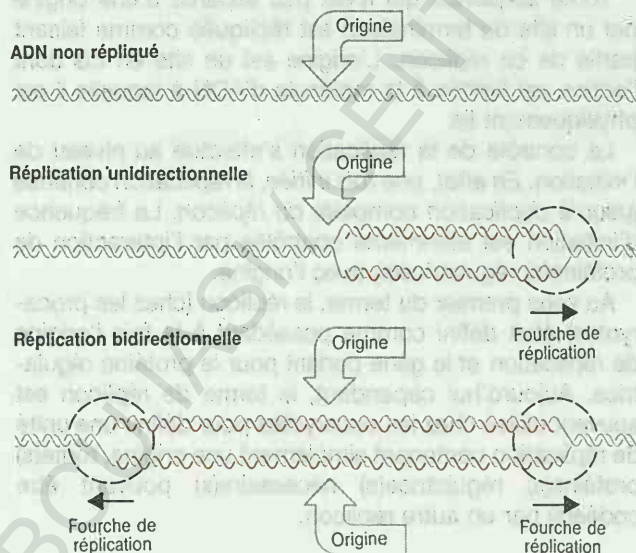
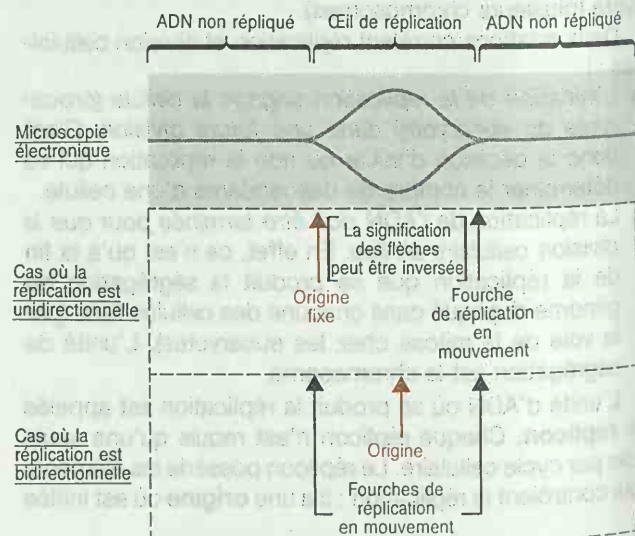


Figure 13.2

Que la réplication soit unidirectionnelle ou bidirectionnelle, il y a formation d'un œil de réplication.



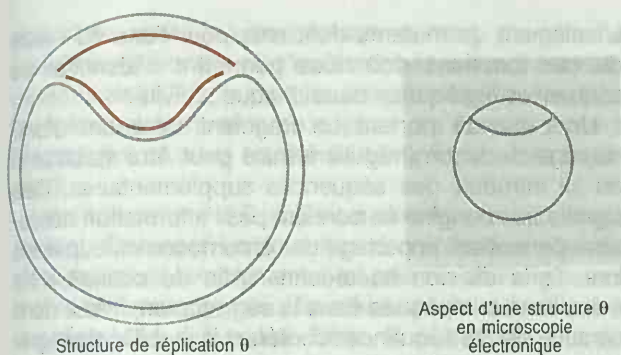


Figure 13.3

Si la molécule d'ADN est circulaire, l'œil de réplication engendre une structure **thêta**.

de **point de croissance**). La fourche de réplication part de l'origine puis avance *séquentiellement* le long de la molécule d'ADN.

La réplication peut être **unidirectionnelle** ou **bidirectionnelle** selon qu'il y a une ou deux fourches de réplication à l'origine. Dans les deux cas, les fourches de réplication partent de l'origine mais si la réplication est bidirectionnelle, les fourches vont ensuite avancer en sens opposé.

En microscopie électronique, il est possible de distinguer l'ADN répliqué qui décrit une forme en **œil** au milieu de l'ADN non répliqué, mais non de savoir s'il s'agit d'une réplication unidirectionnelle ou bidirectionnelle.

En effet, comme le décrit la **Figure 13.2**, l'œil de réplication peut représenter indifféremment l'une ou l'autre des structures. Si la réplication est unidirectionnelle, l'œil est délimité d'une part, par l'origine, d'autre part, par la fourche de réplication. Si la réplication est bidirectionnelle, l'œil est délimité par les deux fourches de réplication qui avancent en sens opposé. Dans les deux cas, l'œil s'étendra jusqu'à réplication complète du réplicon.

Si le réplicon est circulaire, la présence d'un œil confèrera la **structure thêta** dessinée sur la **Figure 13.3**. Les étapes successives de la réplication de l'ADN circulaire du virus polyome ont été visualisées en microscopie électronique sur la **Figure 13.4**.

L'on peut déterminer de deux façons si un œil a une ou deux fourches de réplication. Le choix de la méthode dépend de la molécule d'ADN à laquelle on a affaire, s'il s'agit d'une molécule d'ADN définie ou d'une région inconnue d'ADN du génome cellulaire.

Si la molécule d'ADN est définie et linéaire; la microscopie électronique va nous permettre de mesurer la

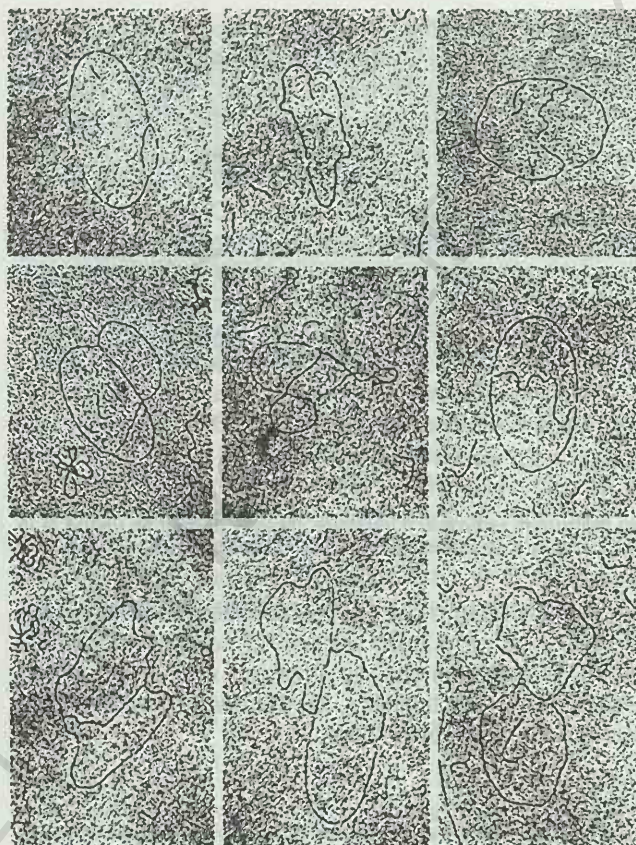


Figure 13.4

Au fur et à mesure que les fourches de réplication avancent le long du réplicon, l'œil de réplication s'agrandit.

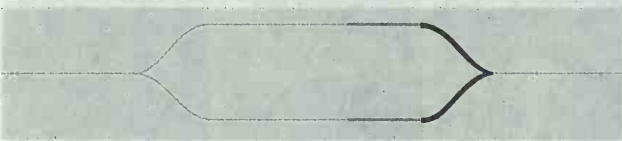
Remarquez comme « l'œil » envahit progressivement l'ADN non répliqué. Il est possible de localiser l'œil de réplication car les deux molécules qui le définissent sont d'égale longueur.

Avec l'aimable autorisation de Bernard Hirt.

distance qui existe entre chaque extrémité de l'œil et chaque extrémité de l'ADN. Ainsi, il est possible de suivre le mouvement de l'œil de réplication d'une molécule d'ADN au cours du temps. Si la réplication est unidirectionnelle, une seule des extrémités de l'œil va se déplacer, l'autre restant fixe à l'origine. Si la réplication est bidirectionnelle, les deux extrémités vont se déplacer, l'origine se trouvant alors à mi-chemin entre elles.

Dans des régions non définies d'un grand génome, le mouvement des fourches de réplication peut être suivi au moyen de deux brefs marquages radioactifs. L'un des marquages est plus intense que l'autre afin qu'ils puissent être distingués par l'intensité relative des marquages obtenus visualisés par autoradiographie. La **Figure 13.5**

Réplication unidirectionnelle



Réplication bidirectionnelle



- Région non marquée radioactivement (invisible sur autoradiographie)
 — Marquage radioactif de faible intensité
 — Marquage radioactif de forte intensité

Figure 13.5

Un marquage radioactif d'intensité variable permet de définir si la réplication est unidirectionnelle ou bidirectionnelle.

montre que dans le cas d'une réplication unidirectionnelle, les deux marquages successifs ne sont décelables qu'à une extrémité de l'œil alors que dans le cas d'une réplication bidirectionnelle, le marquage est identique aux deux extrémités de l'œil.

LE GÉNOME BACTÉRIEN EST UN RÉPLICON UNIQUE

Le génome d'*E. coli* est répliqué de façon bidirectionnelle à partir d'une seule origine, identifiée comme étant le locus génétique *oriC*. L'ADN de l'origine de réplication a la propriété de permettre la réplication de toute séquence d'ADN à laquelle il est lié. Si l'on clone la séquence d'ADN de l'origine de réplication dans une molécule démunie de sa propre origine (mais comportant un marqueur génétique approprié), le recombinant ainsi obtenu est un plasmide capable de réplication autonome *uniquement si l'origine de réplication contient toutes les séquences nécessaires à son identification comme une véritable origine de réplication*. (Une approche similaire a été utilisée pour identifier l'ADN centromérique ou télomérique chez la levure, voir Chapitre 25).

Toute origine pleinement active doit remplir plusieurs fonctions :

- initiation du cycle de réplication ;
- contrôle de la fréquence d'initiation de la réplication ;
- ségrégation des chromosomes répliqués dans les cellules filles.

L'isolement de mutants déficients pour l'une ou l'autre de ces fonctions doit nous permettre d'identifier les séquences impliquées dans chaque activité.

Un plasmide portant un fragment minimum d'*oriC* ségrège de façon irrégulière mais peut être stabilisé si on lui introduit des séquences supplémentaires. Cela signifie que l'origine ne contient pas l'information nécessaire permettant le partage des chromosomes dupliqués lors de la division bactérienne. Afin de connaître les mécanismes impliqués dans la ségrégation, il faut donc caractériser les séquences conférant la stabilité de ségrégation au plasmide.

Des plasmides recombinants comportant environ 1 kb d'ADN provenant de la région d'*oriC* sont maintenus dans *E. coli* au niveau étroitement contrôlé de 1-2 copies par chromosome bactérien. Ainsi, le plasmide a un comportement semblable au chromosome bactérien. En réduisant la taille du fragment d'*oriC* cloné, il a été possible de définir une région fonctionnelle de 245 pb. Cette région assure la réplication du plasmide qui la porte mais ne permet pas le contrôle de la fréquence d'initiation de la réplication : en effet, le plasmide est présent en multiples copies alors que le chromosome bactérien, lui, reste unique.

L'origine de réplication de la bactérie *S. typhimurium* a été localisée dans un fragment d'ADN de 296 pb. La comparaison avec *E. coli* montre qu'il y a 86 % d'homologie de séquence entre les deux origines. Des origines provenant d'autres bactéries peuvent fonctionner dans *E. coli* même si l'homologie de séquence est faible. La comparaison entre différentes origines a permis d'identifier des groupes de courtes séquences conservées entre les différentes bactéries mais pas de savoir si, et comment, ils sont impliqués dans la réplication.

Les séquences des origines de phages lambdaïdes (dont lambda est un exemple) montrent aussi une grande similitude entre elles. En revanche, elles sont bien différentes des origines bactériennes.

Afin d'identifier leurs séquences fonctionnelles, les origines sont modifiées par mutagenèse dirigée puis réintroduites dans les bactéries afin de déterminer les effets de ces modifications. Un grand nombre de mutations ponctuelles n'affectent en rien la capacité du fragment *oriC* de 245 pb à être une origine de réplication. Il est probable qu'il existe dans la région décrite plusieurs sites de courte séquence capables de fixer les protéines régulatrices.

Le chromosome bactérien étant circulaire, cela pose différents problèmes. En effet, les deux fourches de

réplication vont-elles se déplacer le long du génome jusqu'à se rencontrer où s'arrêteront-elles avant à un site de terminaison ? Comment les enzymes impliquées dans la réplication vont-elles se séparer du chromosome à l'endroit précis où la totalité de la réplication est terminée ?

Le chromosome d'*E. coli* a un site de terminaison où cessent d'avancer les fourches de réplication. Ce site n'est pas placé exactement à mi-chemin de l'origine sur le chromosome ; les deux fourches ont donc à parcourir des distances différentes. Chez *B. subtilis*, le site de terminaison n'est pas une structure linéaire mais une fourche d'ADN très sensible aux nucléases. S'il est inséré dans une autre molécule d'ADN, le site de terminaison du plasmide R6K permet l'arrêt de la réplication ; le mécanisme d'action de cette terminaison n'est pas encore connu.

RÉPLICATION ET CYCLE CELLULAIRE

Chez les bactéries, deux liens unissent réplication et croissance cellulaire :

- La fréquence d'initiation de la réplication est donnée par le taux de croissance de la cellule.
- La fin de la réplication est directement liée à la division cellulaire.

La durée de croissance d'*E. coli* peut varier de 18 minutes à 180 minutes. Le chromosome bactérien étant un réplicon unique, la fréquence des cycles de réplication est contrôlée par le nombre d'événements d'initiation à l'origine. Le cycle de réplication peut être défini en terme de deux constantes.

C est un temps donné, d'environ 40 minutes, qui permet la réplication complète du chromosome bactérien. Sa durée est proportionnelle à la vitesse de la fourche de réplication qui parcourt environ 50 000 pb/minute (la vitesse de synthèse de l'ADN est quasi constante à une température donnée et si les précurseurs de la synthèse d'ADN ne sont pas en quantité limitante).

D est un temps donné, d'environ 20 minutes, qui sépare la fin de la réplication de la division cellulaire. C'est le temps requis pour réunir toutes les conditions nécessaires à la division.

(Additionnées, les constantes C et D décrites représentent la vitesse maximum de prolifération que la bactérie

est capable d'adopter ; elles s'appliquent à toutes les vitesses de croissance comprises entre 18 et 60 minutes de temps de doublement, mais si le temps de croissance est supérieur à 60 minutes, la durée de C et D est plus longue.)

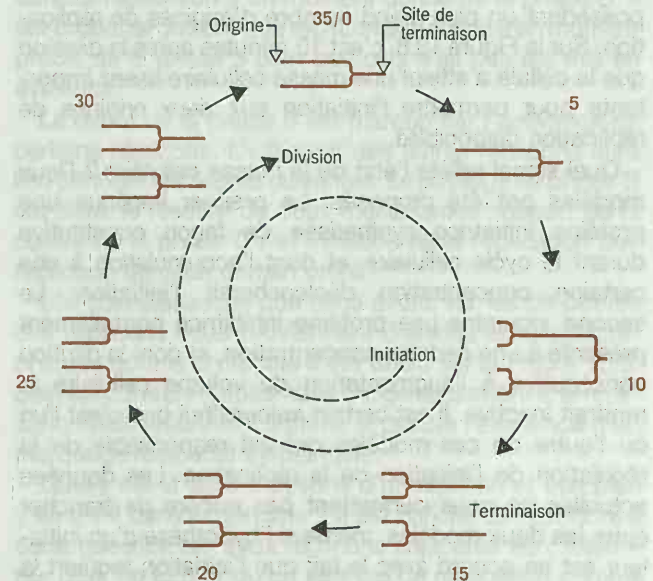
Le cycle de réplication du chromosome doit donc être initié 60 minutes (C + D) avant la division cellulaire. En ce qui concerne les bactéries se divisant plus fréquemment que toutes les 60 minutes, le cycle de réplication doit être initié *avant la fin de la réplication précédente*.

Considérons une bactérie se divisant toutes les 35 minutes. Le cycle de réplication doit être initié 25 minutes avant la division précédente. Cette situation est illustrée sur la **Figure 13.6** où l'état du chromosome entier est schématisé à des intervalles de 5 minutes.

Au moment de la division (35/0 minutes), chaque cellule reçoit un chromosome en cours de réplication dont la fourche de réplication continue d'avancer. A 10 minutes, avant que cette fourche de réplication n'ait atteint le site de terminaison, les origines des chromosomes alors

Figure 13.6

Le temps nécessaire entre l'initiation de la réplication et la division cellulaire étant invariablement de 60 minutes, il y a formation de chromosomes à fourches multiples dans les cellules à croissance rapide.



partiellement répliqués sont initiées. L'initiation de deux nouvelles fourches de réplication donne lieu à un **chromosome « à multiples fourches »**.

A 15 minutes — c'est-à-dire 20 minutes avant la prochaine division — la première fourche de réplication atteint le site de terminaison, permettant la séparation des deux chromosomes fils ; chacun d'entre eux est déjà partiellement engagé dans la réplication grâce aux « nouvelles » fourches de réplication initiées au temps 10 minutes et qui continuent d'avancer.

Au moment de la division, les deux chromosomes partiellement répliqués ségrègent, recréant ainsi la situation initiale. La fourche de réplication atteindra ensuite le site de terminaison à 15 minutes et, 20 minutes plus tard se produira la division. Nous voyons donc que l'initiation de la réplication se produit tous les 25/35° de cycle avant la division suivante.

Le principe général liant initiation et cycle cellulaire est le suivant : *plus la cellule croît rapidement (ou plus le cycle cellulaire est court), plus nombreux sont les cycles de réplication initiés avant chaque division*, donc plus nombreux sont les chromosomes en cours de duplication. En effet, les temps C et D ne pouvant être écourtés, il faut que la cellule prévoit les divisions ultérieures et donc qu'elle initie la réplication 60 minutes avant chaque division même si celle-ci est à plusieurs générations d'intervalle.

Comment la cellule sait-elle à quel moment le cycle de réplication doit être initié ? L'initiation est déclenchée relativement à la masse cellulaire. Les cellules à croissance rapide ont une masse cellulaire plus importante et possèdent un plus grand nombre d'origines de réplication. Sur la Figure 13.6, c'est 10 minutes après la division que la cellule a atteint une masse cellulaire assez importante pour permettre l'initiation aux deux origines de réplication disponibles.

Quel signal reflète l'état de la masse cellulaire ? Deux modèles ont été proposés. Le premier implique une protéine initiateur synthétisée de façon constitutive durant le cycle cellulaire, et dont l'accumulation à une certaine concentration déclencherait l'initiation. Le second incrimine une protéine inhibitrice normalement présente à une certaine concentration, et dont la dilution consécutive à l'augmentation du volume cellulaire la rendrait inactive. Il est certain aujourd'hui que c'est l'un ou l'autre de ces modèles qui est responsable de la régulation de l'initiation de la réplication. Les données actuelles ne nous permettent pas encore de trancher entre les deux modèles, même si l'hypothèse d'un initiateur est en accord avec le fait que l'initiation requiert la synthèse protéique.

Parmi les gènes nécessaires à la division cellulaire, le gène *ftsZ* joue un rôle important. Des mutations dans *ftsZ* bloquent la formation du septum et génèrent des filaments multinucléés. D'autre part, le produit du gène *FtsZ* est la cible d'une autre protéine (*SulA*) qui bloque la division cellulaire dans certaines circonstances. La surexpression de *ftsZ* induit la formation de *minicellules* morphologiquement intactes mais sans ADN. Il semble donc que la fréquence de division soit déterminée par la quantité de protéines *FtsZ*. *FtsZ* a un rôle de protéine initiateur de la division cellulaire.

La croissance bactérienne peut être mesurée en terme d'**unité cellulaire** de 1,7 μm de longueur ; chaque unité cellulaire ayant une origine de réplication. Ainsi, une cellule à croissance rapide avec deux origines de réplication aura une longueur inférieure à 3,4 μm . Le point de croissance est défini topologiquement comme étant le lieu où il y a initiation et où se produit un changement de structure de la cellule. Il y a un point de croissance par unité cellulaire.

Depuis longtemps, on pense qu'il existe un lien physique entre l'ADN et la membrane bactérienne. En effet, dans les fractions membranaires, on trouve de l'ADN bactérien riche en marqueurs génétiques proches de l'origine, de la fourche de réplication et du site de terminaison. Le point de croissance pourrait être une structure membranaire sur laquelle l'origine serait attachée au moment de l'initiation. Chez les mammifères, les fourches de réplication seraient liées à la membrane nucléaire.

Cette liaison entre l'ADN et la membrane bactérienne pourrait aussi expliquer la ségrégation. Si les chromosomes fils sont attachés à la membrane, ils pourraient être physiquement séparés au fur et à mesure que la membrane croît entre eux. La **Figure 13.7** montre qu'ensuite, un septum pourrait se former entraînant la ségrégation de chaque chromosome dans une cellule fille différente. Si nous nous rappelons l'événement de ségrégation tel qu'il est schématisé sur la Figure 13.6, nous nous attendons à ce que le site de terminaison de la réplication sur le chromosome soit impliqué dans ce phénomène.

LES CHROMOSOMES EUCARYOTES CONTIENNENT DE NOMBREUX RÉPLICONS

Dans les cellules eucaryotes, la réplication de l'ADN n'est effectuée que durant une partie du cycle cellulaire. La période qui sépare deux mitoses (interphase) est divisée en différentes phases comme l'indique la **Figure 13.8**.

La cellule commence un cycle cellulaire à l'état diploïde et le reste pendant la période **G1** (qui comprend le plus

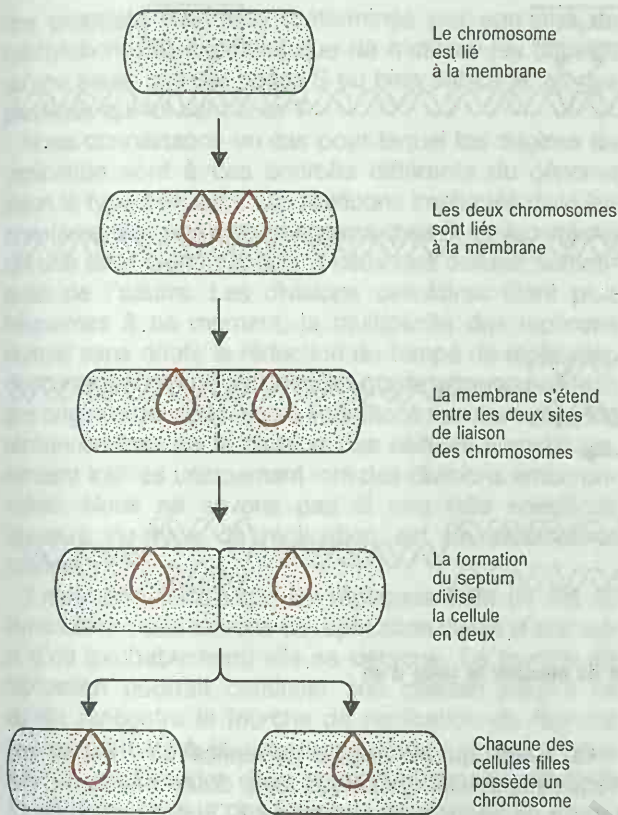


Figure 13.7
La ségrégation des chromosomes est peut-être due à la liaison de l'ADN bactérien à la membrane.

souvent la plus grande partie du cycle mais dont la durée est hautement variable selon les phénotypes).

C'est durant la **phase S** qu'est synthétisé l'ADN. Cette période dure quelques heures chez les eucaryotes supérieurs.

À la fin de la phase S, la cellule est tétraploïde et le restera durant toute la phase **G2**. Durant la mitose (M), un nombre diploïde de chromosomes va être réparti dans chacune des cellules filles (les périodes précédant et succédant la phase S sont appelées G1 et G2 car elles représentent des étapes intermédiaires ou « gaps » dans la synthèse de l'ADN).

Pendant la phase G1, se déroulent les événements responsables de l'initiation de la phase S. L'on sait que la synthèse protéique est nécessaire à cette initiation mais la nature exacte des événements reste obscure. En particulier, on a beaucoup discuté pour savoir si les signaux impliqués étaient cumulatifs ou soudains. A vrai

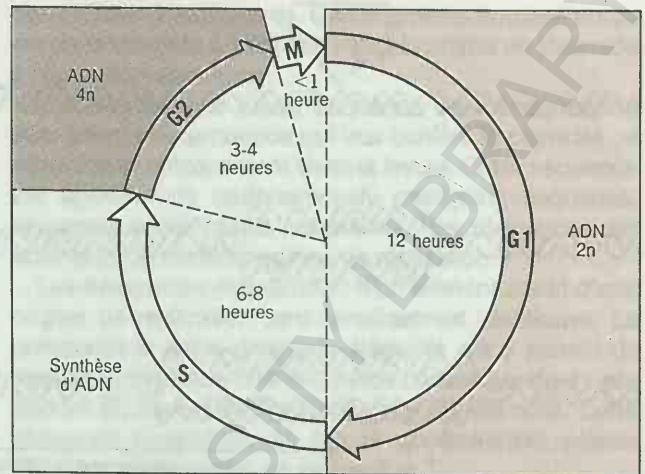


Figure 13.8
Le cycle cellulaire des eucaryotes est divisé en quatre phases : G1, S, G2 et M. Les durées indiquées sont les durées moyennes observées pour des lignées cellulaires de mammifère en culture.

dire, l'on sait très peu de choses au plan moléculaire sur la manière dont la cellule prend la décision de permettre à son génome de se répliquer.

La réplication de la totalité de l'ADN d'un chromosome eucaryote est effectuée grâce à un grand nombre de réplicons. Seule une partie de ces réplicons est engagée dans la réplication à un moment donné de la phase S. Il est probable que chaque réplicon soit activé à un moment précis de la phase S bien que cela n'ait pas été mis en évidence.

Le début de la phase S est marqué par l'activation de certains réplicons. L'initiation des autres réplicons aura lieu pendant les heures suivantes. Le passage en phase S est donc le résultat de deux mécanismes : départ de la cellule de la phase G1 et initiation de la réplication des réplicons dans un certain ordre.

Les connaissances que nous avons sur les propriétés des réplicons nous proviennent d'études autoradiographiques dont le protocole a été décrit sur la Figure 13.5. Les réplicons chromosomiaux ont généralement une réplication bidirectionnelle comme le prouve l'existence d'un marquage radioactif symétrique.

Quelle est la taille moyenne d'un réplicon et combien y en a-t-il dans le génome ? Il est difficile de répondre à cette question car deux réplicons adjacents vont fusionner pour ne former qu'un seul œil de réplication comme l'illustre la Figure 13.9. La seule façon de mesurer la taille

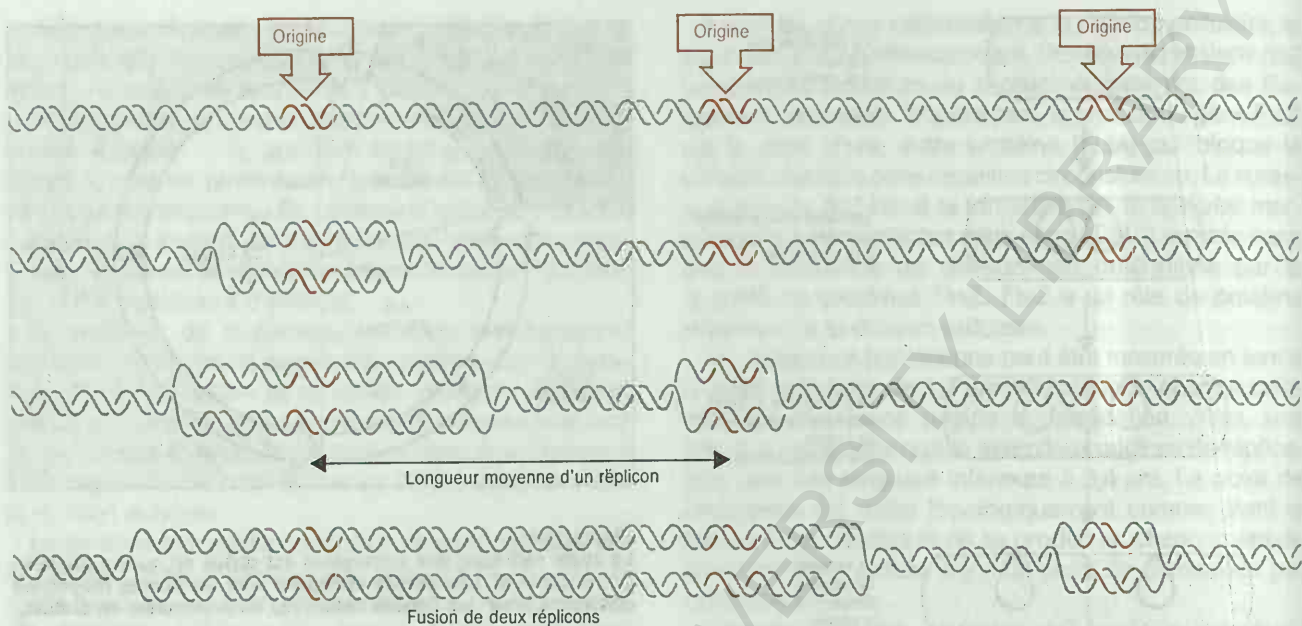


Figure 13.9
Une région d'ADN où l'on peut distinguer deux réplicons actifs permet de mesurer la taille d'un réplicon.

d'un réplicon est de considérer une molécule d'ADN où plusieurs yeux de réplication sont visibles, et de supposer que chaque œil correspond à un réplicon initié au même moment que son voisin mais dont la fourche de réplication est encore indépendante.

(Il y a de sérieux arguments pour penser que des contrôles régionaux pourraient être responsables de ce type d'activation dans lequel des groupes de réplicons sont activés de manière plus ou moins coordonnée ; le mécanisme d'activation de réplicons isolés dispersés dans le génome devrait être différent.)

Dans un groupe de réplicons actifs, la taille moyenne d'une unité de réplication est définie comme étant la distance moyenne séparant deux origines (c'est-à-dire la distance moyenne séparant les points médians de deux réplicons adjacents). La mesure de la longueur marquée radioactivement durant un temps donné puis révélée par autoradiographie, permet d'évaluer la vitesse à laquelle se déplace la fourche de réplication.

Dans le **Tableau 13.1** sont comparés quelques réplicons eucaryotes et procaryotes. Les réplicons eucaryotes sont de petite taille (bien que selon les espèces, les tailles peuvent varier d'un facteur 10) ; leurs fourches de réplication se déplacent bien plus lentement que chez les procaryotes.

Qu'est-ce qu'une origine de réplication ? Est-ce une séquence d'ADN spécifique sans doute membre d'une famille de séquences répétitives ? Ou est-ce une structure adoptée par la fibre chromatinienne pas obligatoirement associée à un type unique de séquence ?

Quel est le mécanisme qui va permettre à telle origine d'être initiée à tel moment durant la phase S ? Comment sont distinguées les origines ayant déjà été répliquées de celles qui n'ont pas encore été répliquées ? Est-ce

Tableau 13.1
Les réplicons eucaryotes sont plus petits et se répliquent moins vite que le réplicon bactérien.

Organisme	Nombre de réplicons	Longueur moyenne	Vitesse de la fourche de réplication
<i>E. coli</i>	1	4 200 kb	50 000 pb/min
<i>S. cerevisiae</i>	500	40 kb	
<i>D. melanogaster</i>	3 500	40 kb	2 600 pb/min
<i>X. laevis</i>	15 000	200 kb	500 pb/min
<i>M. musculus</i>	25 000	150 kb	2 200 pb/min
<i>V. faba</i>	35 000	300 kb	

une propriété de l'ADN (déterminée par son état de méthylation, par exemple) que de n'utiliser les origines qu'une seule fois par phase S ou bien est-ce le fait des protéines qui lui sont liées ?

Nous connaissons un cas pour lequel les origines de réplication sont à des endroits différents du génome selon le type cellulaire. Les réplicons impliqués dans les premières divisions embryonnaires chez *D. mélagonaster* ont une taille bien inférieure à ceux des cellules somatiques de l'adulte. Les divisions cellulaires étant plus fréquentes à ce moment, la multiplicité des réplicons permet sans doute la réduction du temps de réplication des chromosomes. Il est possible que le génome possède des origines de réplication à spécificité tissulaire qui, non reconnues lors de la division des cellules somatiques, seraient initiées uniquement lors des divisions embryonnaires. Nous ne savons pas si une telle spécificité tissulaire du type de réplication est un phénomène habituel.

Il n'est pas certain que les réplicons aient un site de terminaison où la fourche de réplication cesse d'avancer et d'où (probablement) elle se détache. La fourche de réplication pourrait continuer son chemin jusqu'à ce qu'elle rencontre la fourche de réplication du réplicon adjacent. Il reste à élucider le problème topologique que pose la jonction des deux brins d'ADN néosynthétisés au point de jonction des fourches de réplication.

ISOLEMENT DES ORIGINES DES RÉPLICONS CHEZ LA LEVURE

Tout segment d'ADN possédant une origine de réplication doit être capable de se répliquer. Bien qu'il soit rare de trouver des plasmides chez les eucaryotes, il est possible de construire de telles molécules par manipulation génétique.

L'œuf de *Xénopus* est un système qui permet la réplication de molécules d'ADN extrachromosomiques. Si l'on injecte une molécule d'ADN (circulaire), elle ne se réplique qu'une seule fois par cycle cellulaire et ceci quelle que soit la molécule. Dans ce système, il semble donc que la présence d'une origine ne soit pas nécessaire.

Un autre système permet d'isoler des origines de réplication. Des cellules de levure *S. cerevisiae*, mutées pour certaines fonctions, peuvent être « transformées » après introduction de séquences d'ADN possédant une copie sauvage du gène muté. Des fragments d'ADN de levure sont capables de transformer à très haute efficacité

des cellules défectueuses. Ces fragments peuvent survivre dans la cellule à l'état non intégré comme un plasmide à réplication autonome.

Les fragments à haute fréquence de transformation possèdent une séquence qui leur confère la capacité de se répliquer efficacement dans la levure. Cette séquence est appelée *ars* (autonomously replicating sequence, séquence à réplication autonome). Les séquences *ars* sont-elles de réelles origines de réplication ?

Les fréquences d'apparition d'un élément *ars* et d'une origine de réplication sont sensiblement identiques. La comparaison entre différents éléments *ars* a permis de mettre en évidence une séquence consensus de 11 pb, portant au maximum 3 substitutions par élément. Cette séquence consensus se trouve généralement placée dans une région riche en bases A et T.

Par mutations ponctuelles, il a été défini dans l'élément *ars* une région « cœur » de 14 pb qui confère la fonction d'origine de réplication. La région cœur comprend les 11 pb de la séquence consensus précédemment décrite. Il est possible que la séquence consensus soit le site de reconnaissance pour la fixation d'une protéine participant à la fonction d'origine de réplication ; si d'autres séquences sont impliquées, elles sont nécessairement différentes d'un élément *ars* à l'autre.

Les origines de réplication chez la levure ne sont pas forcément membres d'une famille de séquences répétitives. Il est possible que différentes séquences (n'ayant en commun que la courte séquence consensus) aient la fonction d'origine de réplication.

DIFFÉRENTES STRUCTURES DE RÉPLICATION : ŒIL DE RÉPLICATION, CERCLE ROULANT ET BOUCLE D

La réplication donne lieu à différentes structures déterminées par deux paramètres : selon que le modèle d'ADN à copier est circulaire ou linéaire et selon que la fourche de réplication est engagée dans la synthèse de l'un ou des deux brins d'ADN.

Considérons la fourche de réplication comme étant le lieu de synthèse des deux nouveaux brins d'ADN. Dans le cas d'une molécule d'ADN linéaire, le mouvement de la (ou des) fourche(s) de réplication (selon que la réplication est unidirectionnelle ou bidirectionnelle) donne lieu à un œil de réplication (Figure 13.2). Par contre, si la molécule est circulaire, la réplication engendre une structure thêta (Figure 13.3)

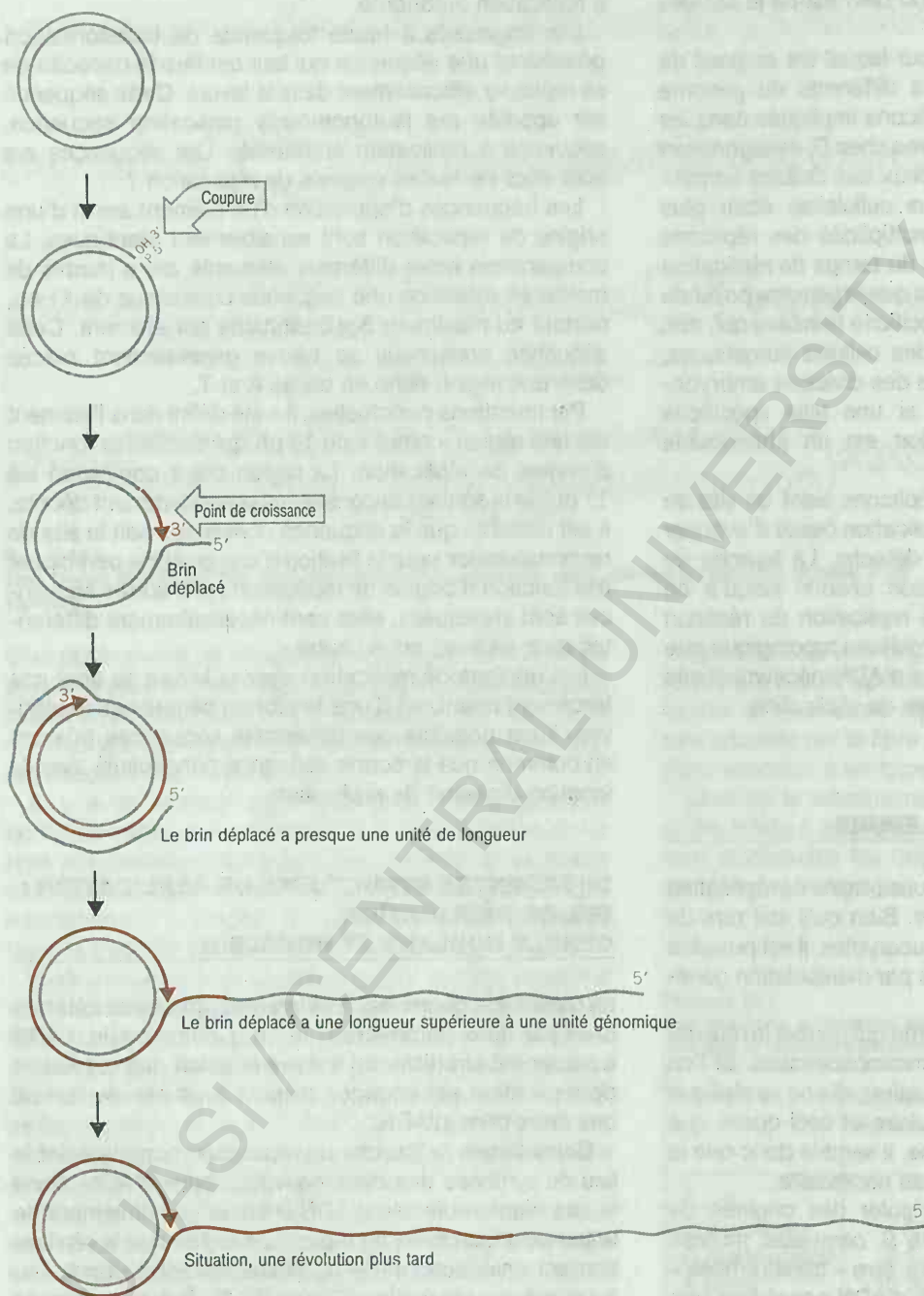


Figure 13.10

La réplication par le cercle roulant engendre une molécule monocaténaire multimérique (qui peut être convertie en une molécule bicaténaire grâce à la synthèse du brin complémentaire ; non décrit sur la Figure).

Supposons maintenant qu'à l'origine, ne soit initiée la réplication que d'un seul des deux brins d'ADN. Dans ce cas, la réplication d'une molécule circulaire produit une nouvelle structure. La coupure d'un des brins d'ADN engendre une extrémité 3'-OH libre qui va être allongée par l'ADN polymérase. Ainsi, le brin nouvellement synthétisé déplace progressivement le brin parental. Les différentes étapes de ce modèle de réplication sont schématisées sur la **Figure 13.10**.

Ce type de structure est appelé **cercle roulant**, car le point de croissance de la réplication « roule », en quelque sorte, le long de la matrice d'ADN circulaire. En principe, cela pourrait continuer indéfiniment. Au cours de son mouvement, la fourche de réplication allonge le brin externe et déplace le brin synthétisé précédemment.

Le brin nouvellement synthétisé étant lié de façon covalente au brin externe initial, la chaîne en cours de réplication possède à son extrémité 5' libre l'unité génomique originelle (complémentaire de la matrice d'ADN) suivie d'un certain nombre de copies néosynthétisées. Le nombre de ces unités génomiques est égal au nombre de révolutions effectuées autour de la matrice, chaque révolution déplaçant la totalité du matériel synthétisé lors du cycle précédent.

Un exemple de cercle roulant a été photographié en

microscopie électronique sur la **Figure 13.11**. Le cercle roulant est utilisé à différents desseins *in vivo*.

Le génome de certains phages ne consiste qu'en une molécule circulaire d'ADN simple brin. Si leur réplication se fait suivant le mode du cercle roulant, ils doivent en premier lieu convertir leur génome en une molécule circulaire double brin. Les copies du génome nouvellement synthétisées (et donc sous forme simple brin) vont ensuite être clivées les unes des autres. Elles seront alors insérées dans des particules de phages ou utilisées pour de futurs cycles de réplication. (Ce système de réplication est discuté dans le Chapitre 14.)

Les molécules simple brin engendrées par le cercle roulant peuvent être converties en molécules double brin par synthèse du brin complémentaire. Il y a alors formation d'une molécule linéaire double brin constituée de copies génomiques en tandem. Ces molécules sont impliquées dans la maturation de certains phages, comme le phage lambda (voir Chapitre 25).

Le mécanisme du cercle roulant est utilisé dans l'oocyte de *Xenopus* pour amplifier l'ADNr. Une unité génomique d'ADNr prend la forme d'un cercle roulant. Le brin multicopies néosynthétisé est transformé en une molécule double brin. La molécule est ensuite séparée du cercle par clivage ; la jonction de ses deux extrémités donne lieu à une molécule circulaire contenant un grand nombre de copies d'ADNr. Il y a donc eu amplification de l'ADNr.

Une autre forme de réplication a été décrite dans certaines mitochondries. La molécule d'ADN circulaire double brin a toujours une origine d'où part la fourche de réplication, mais seul l'un des deux brins parentaux est copié (le brin H de l'ADN mitochondrial chez les mammifères). La synthèse du nouveau brin déplace progressivement l'ADN complémentaire (brin L) qui reste simple brin comme l'illustre la **Figure 13.12**. Cette structure de réplication est appelée **boucle D**, D pour **déplacement**.

L'ouverture des deux brins se fait sur une région de 500 à 600 bases dans les mitochondries de mammifères. La région du brin L où se produit l'ouverture est instable. Elle est souvent dégradée puis resynthétisée.

Chez *X. laevis*, l'ouverture des deux brins de l'ADN mitochondrial se produit sur une région plus longue. Il peut exister plusieurs boucles D sur une même molécule ; sur la molécule d'ADN linéaire des mitochondries de *Tetrahymena*, on en a dénombré six. L'ADN des chloroplastes se réplique selon le même mécanisme, mais avec deux boucles D (chez les plantes supérieures).

La synthèse du brin complémentaire de l'ADN mitochondrial se fait par extension du brin court situé dans

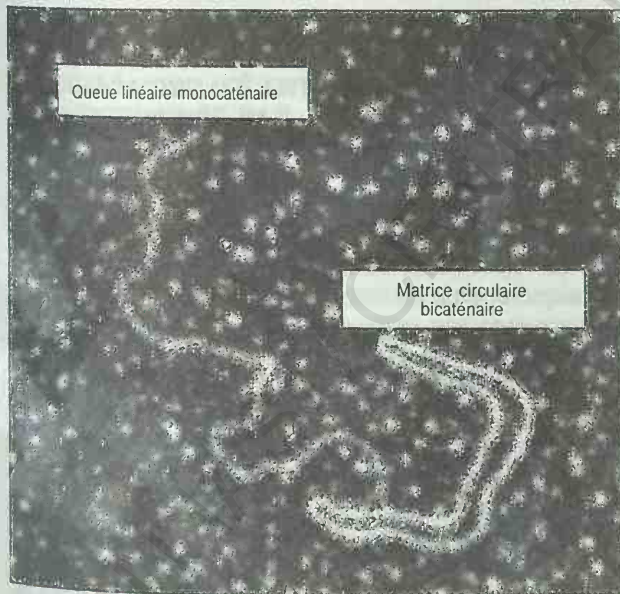


Figure 13.11

En microscopie électronique, une molécule dont la réplication s'effectue selon un cercle roulant apparaît comme une molécule circulaire possédant une queue linéaire.

Avec l'aimable autorisation de David Dressler.

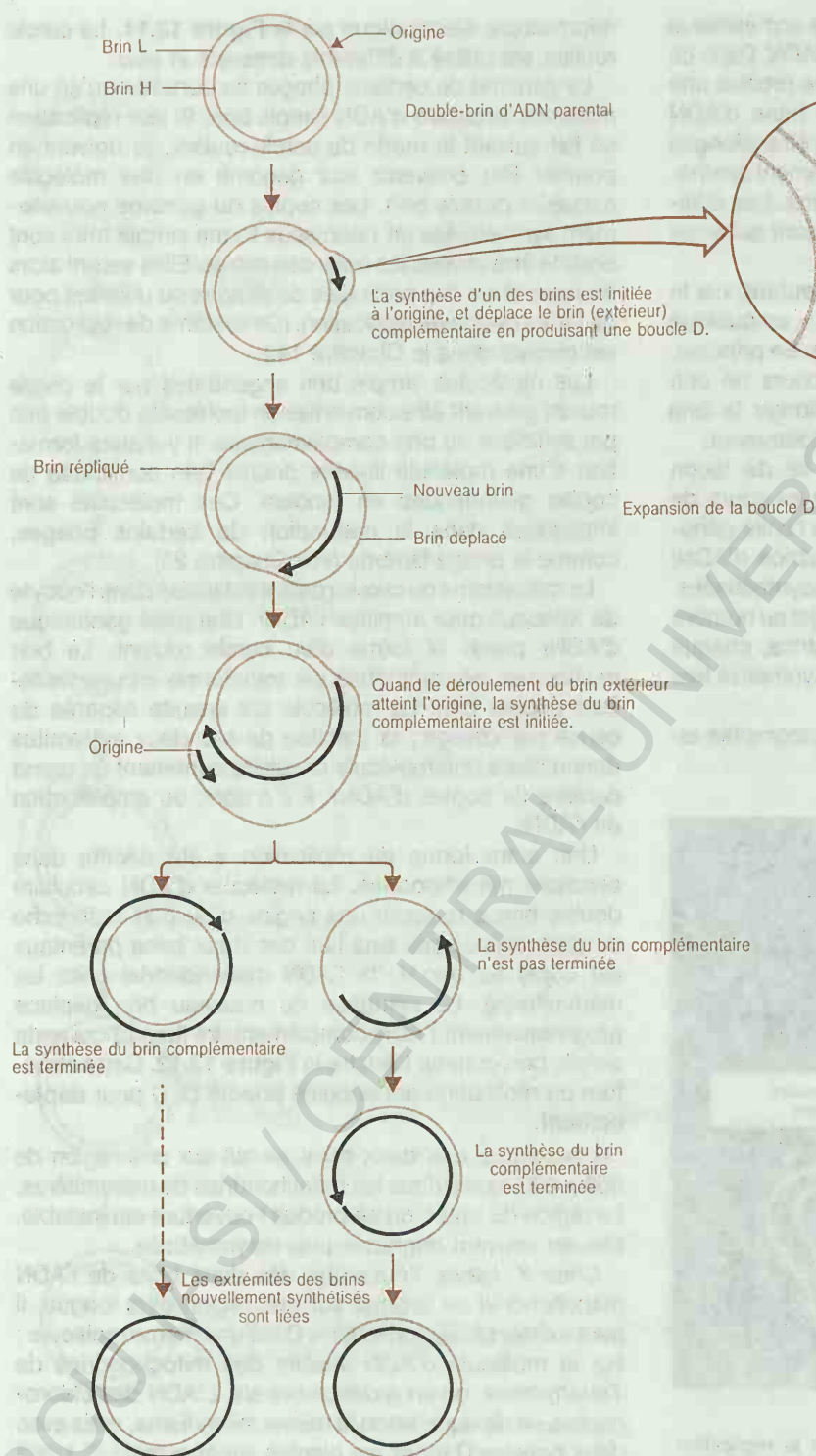


Figure 13.12

La boucle D permet de maintenir une ouverture dans les molécules d'ADN mitochondrial de mammifère dont la réplication s'effectue à partir d'une origine différente pour chacun des brins.

la boucle D. Le brin L est déplacé au fur et à mesure que la boucle D s'étend. Aux deux tiers de la synthèse, lorsque l'origine de réplication du brin L devient monocaténaire, la synthèse d'un brin H est initiée. Elle se poursuivra le long du brin L, en sens opposé à la synthèse du nouveau brin L déjà aux deux tiers formé.

A cause du décalage qui existe entre l'initiation des deux origines de réplication, un tiers seulement du nouveau brin H a été synthétisé quand est achevée la synthèse du brin L. Il y a libération, d'une part, d'une molécule circulaire totalement double brin, et d'autre part d'une molécule partiellement simple brin jusqu'à ce que la synthèse du brin H soit terminée. Les extrémités des molécules nouvellement synthétisées sont ensuite reliées entre elles pour former des molécules circulaires intactes.

L'existence même de cercles roulants et de boucles D est donc la preuve qu'une origine de réplication peut être une séquence servant à l'initiation de la synthèse d'un seul des deux brins d'ADN. L'ouverture de la molécule d'ADN double brin ne conduit pas nécessairement à la réplication des deux brins, chaque brin pouvant avoir sa propre origine de réplication.

Le corollaire est qu'une origine initiant la réplication des deux brins d'ADN le fait en vertu d'une configuration particulière : les origines de réplication des deux brins sont situées en vis-à-vis au même point. De cette configuration dépendra aussi l'uni- ou bidirectionnalité de la réplication.

L'INCOMPATIBILITÉ ENTRE PLASMIDES EST DÉTERMINÉE PAR LE NOMBRE DE COPIES

Chaque type de plasmide est maintenu à un certain nombre de copies dans son hôte bactérien. Certains plasmides partagent le nombre de réplifications de leur hôte bactérien : ce sont des plasmides **simple-copie** qui sont présents en un seul exemplaire par chromosome bactérien. D'autres sont **multicopies** et sont présents en un nombre qui leur est propre (généralement 10-20) par chromosome bactérien.

Le nombre de copies est directement déterminé par le mécanisme qui contrôle la réplication. A chaque initiation de la réplication est associé un certain nombre d'origines par bactérie. Comme chaque plasmide est un réplicon unique, le nombre d'origines est le même que le nombre de molécules plasmidiques.

Les plasmides simple-copie ont un système de contrôle

similaire à celui qui gouverne la réplication du chromosome bactérien. Une origine ne peut être initiée qu'une seule fois ; les origines filles ségrègent ensuite dans les deux cellules filles.

Les plasmides multicopies ont un système de réplication qui permet la coexistence de plusieurs origines déterminant le nombre total de copies. Si le plasmide est amené à survivre dans les bactéries, il doit aussi exister un système qui assure une bonne ségrégation des plasmides dans les cellules filles.

L'**incompatibilité entre plasmides** est une conséquence de la régulation du nombre de copies. Un **groupe de compatibilité** est défini comme étant un groupe de plasmides incapables de coexister dans la même bactérie. Incapables de se distinguer entre eux, ils mettent en péril les étapes essentielles à leur survie (comme la réplication et la ségrégation).

Un modèle de contrôle négatif, imaginé pour expliquer l'incompatibilité entre plasmides, implique un répresseur qui serait capable de mesurer le nombre de copies par mesure du nombre d'origines (ce modèle est identique à celui décrit dans la régulation de la réplication bactérienne).

L'apparition d'une nouvelle origine de réplication, due à l'introduction d'un second plasmide du même groupe de compatibilité, a pour effet de simuler la réplication du plasmide résident : il y a maintenant deux origines. Aucune réplication ne sera initiée avant que les deux plasmides ne ségrègent dans des cellules différentes, créant ainsi la situation de préréplication comme l'illustre la **Figure 13.13**.

Si le système permettant la ségrégation dans les cellules filles ne peut faire la distinction entre les deux plasmides, il y a aussi incompatibilité. Les plasmides ségrègent dans des cellules différentes et, par conséquent, ne survivent pas ensemble dans une lignée cellulaire.

En termes simples, le plasmide est égoïste. Une fois qu'il a envahi une bactérie, il va tout faire pour en empêcher l'accès à un autre plasmide du même groupe de compatibilité. L'incompatibilité est un moyen pour le plasmide d'établir ses droits territoriaux.

La présence d'un plasmide appartenant à un autre groupe de compatibilité n'affecte pas directement la survie du plasmide résident. En effet, bien qu'il ne puisse survivre qu'un seul réplicon par groupe de compatibilité dans une bactérie, des réplicons de différents groupes de compatibilité peuvent coexister dans une même cellule

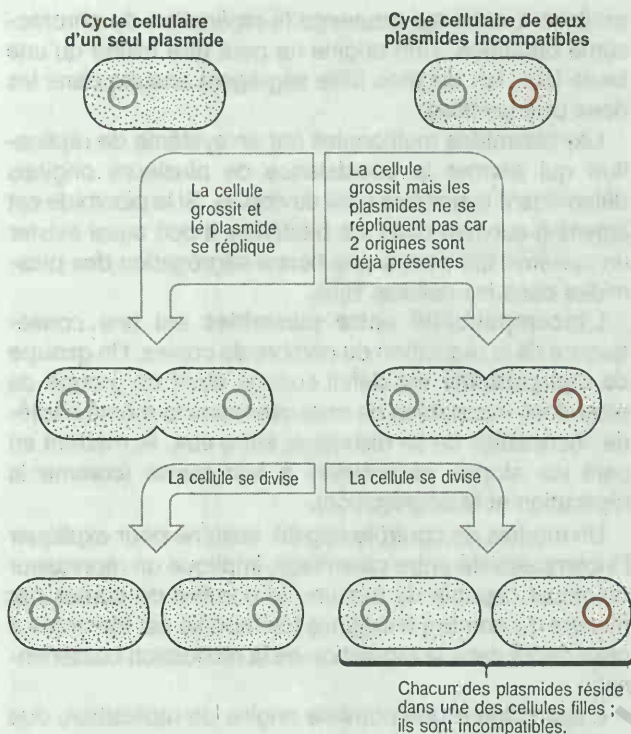


Figure 13.13

Deux plasmides sont incompatibles (ils appartiennent au même groupe de compatibilité) quand l'origine de l'un ne peut être distinguée de l'origine de l'autre au moment de l'initiation. Ce modèle est également valable si c'est le site nécessaire à la répartition dans les cellules filles qui n'est pas reconnu au moment de la ségrégation.

(il n'est pas exclu, qu'en conditions limitantes, les deux plasmides entrent en compétition pour assurer leur survie).

Le système qui a été le mieux étudié est celui du plasmide ColE1 de *E. coli*. Avant d'entrer dans les détails d'un tel système, il est nécessaire de décrire brièvement le mécanisme d'initiation de la réplication de l'origine de ColE1, dont les différentes étapes sont schématisées sur la Figure 13.14.

La réplication commence par l'initiation de la transcription d'un ARN à 555 pb en amont de l'origine. Cette espèce est dénommée « ARN amorce ». La transcription se poursuit au-delà de l'origine. L'enzyme ARNase H (H car son substrat est un ARN *hybridé* à une molécule d'ADN) coupe le transcrit à l'endroit où il rencontre l'origine. Cela engendre une extrémité 3'-OH libre qui est utilisée comme amorce à partir de laquelle la synthèse d'ADN est initiée.

L'ARN « amorce » subit deux types de régulation, l'un

par un ARN complémentaire, l'autre par une protéine codée par un locus proche.

L'ARN I régulateur est une molécule de 108 bases, synthétisée à partir du brin complémentaire du brin d'ADN permettant la transcription de l'ARN amorce. Les positions relatives de l'ARN I et de l'ARN amorce sont décrites sur la Figure 13.15. La synthèse de l'ARN I est initiée au cœur de la région de 555 pb et se termine au niveau où l'ARN amorce a été initié. L'ARN I est donc complémentaire de la région 5' terminale de l'ARN amorce. Une molécule d'ARN qui fonctionne grâce à sa complémentarité avec un autre ARN codé par une même région d'ADN est appelée **contre-transcrit**.

Il est possible de réduire, voire de supprimer l'incompatibilité entre plasmides en mutant des plasmides du même groupe de compatibilité, puis en les sélectionnant pour leur capacité à coexister dans une même bactérie. Les mutations dans la région commune à l'ARN I et à l'ARN amorce de ColE1 réduisent son incompatibilité avec d'autres plasmides du même groupe de compatibilité. Il semble donc que l'un des ARN, si ce n'est les deux, soit impliqué dans le mécanisme d'incompatibilité.

Si l'on ajoute l'ARN I dans un système de réplication de ColE1 *in vitro*, il y a inhibition de la formation de l'ARN amorce. L'ARN I n'inhibant ni l'initiation, ni l'élongation de la transcription de l'ARN amorce, sa présence doit empêcher l'ARNase H de fonctionner et d'engendrer l'extrémité 3'-OH de l'ARN amorce. Ceci est probablement le résultat de l'appariement des deux ARN entre eux.

Les deux molécules ont la même structure secondaire, c'est-à-dire qu'elles possèdent trois structures en épingle à cheveux dont la boucle terminale est simple brin. Les mutations réduisant l'incompatibilité entre plasmides sont localisées dans ces boucles simple brin, ce qui suggère que l'appariement entre les deux ARN se fait grâce au contact entre les boucles.

Comment l'appariement avec l'ARN I prévient-il le clivage qui engendre l'amorce ? Un modèle est présenté dans la Figure 13.16. En l'absence de l'ARN I, l'ARN amorce adopte une structure secondaire telle que sa région 5' terminale est hybridée avec une autre partie de la molécule. S'il est présent, l'ARN I s'hybride sur toute sa longueur avec l'extrémité 5' terminale de l'ARN amorce, modifiant ainsi la structure secondaire de la molécule. Cette nouvelle structure ne permet pas le clivage de l'ARN amorce par l'ARNase H.

Ce modèle rappelle quelque peu le mécanisme impliqué dans l'atténuation de la transcription pour laquelle il est clair que la présence d'une séquence d'ARN permet ou empêche la formation de la structure secondaire nécessaire à la terminaison par l'ARN polymérase, selon

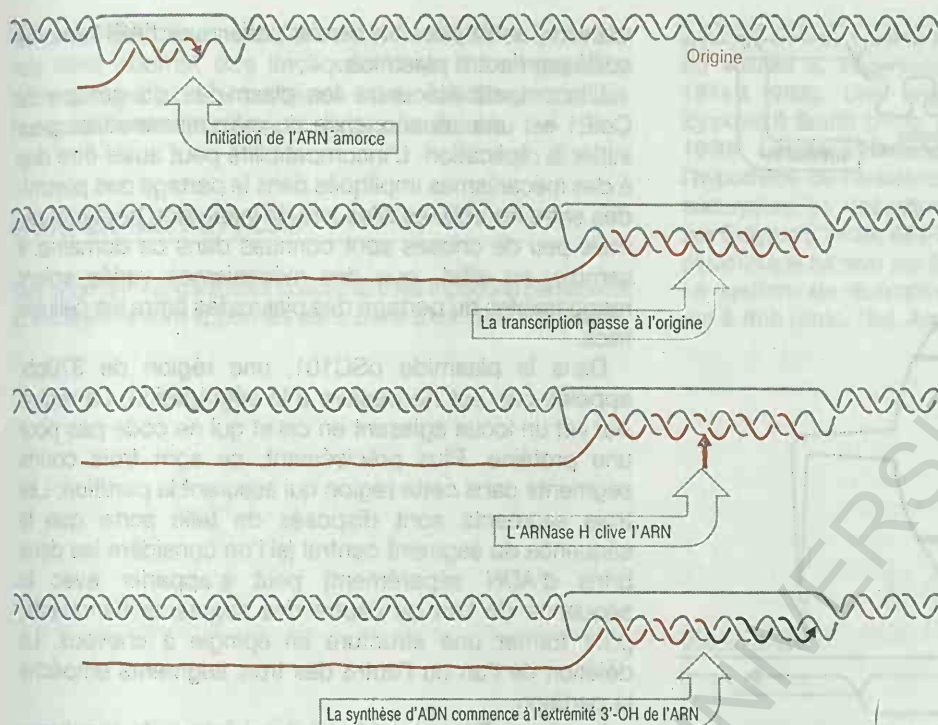


Figure 13.14

La réplication de l'ADN de ColE1 est initiée grâce au clivage de l'ARN amorce qui engendre une extrémité 3'-OH. L'appariement de l'ARN I avec l'amorce empêche le clivage.

(Nous ne connaissons pas les bases de cet effet inhibiteur. L'action de l'ARNase H est peut-être dépendante de la reconnaissance de molécules hybrides ADN/ARN et non pas de la structure de l'ARN.)

qu'il y a ou non appariement (voir Chapitre 11). L'action de l'ARN I est donc due à sa capacité à altérer la structure secondaire d'une région située à distance de l'amorce.

Formellement, ce modèle revient à postuler l'existence d'une boucle de contrôle faisant intervenir deux espèces d'ARN : l'ARN amorce, un régulateur positif nécessaire à l'initiation de la transcription, et l'ARN I, un régulateur négatif capable d'inhiber l'action du régulateur positif.

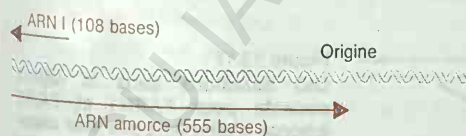


Figure 13.15

La séquence de l'ARN I est complémentaire de la région 5' de l'ARN amorce.

L'ARN I est capable d'agir sur la réplication de tout plasmide présent dans la cellule. Tout comme le répresseur lysogénique de lambda (voir Chapitre 12), l'ARN I agit comme répresseur, empêchant le fonctionnement de tout nouvel arrivant dans la cellule. Au lieu d'un répresseur protéique se liant à la nouvelle molécule d'ADN, c'est un ARN se liant à un précurseur nouvellement synthétisé, l'ARN amorce, qui assure le contrôle de la réplication.

Une petite protéine (63 acides aminés) est codée par le gène *rop*, situé en aval de l'origine. La protéine Rop permet l'appariement de l'ARN I et de l'ARN amorce et, par conséquent, inhibe la formation de l'amorce.

En quoi une mutation dans la séquence des ARN peut-elle affecter l'incompatibilité entre plasmides ? Une mutation dans la séquence codant pour les deux ARN, chez l'un ou l'autre des plasmides, peut empêcher l'appariement entre ARN I et l'ARN amorce, comme le montre la Figure 13.17. Chaque ARN I continue de s'apparier avec l'ARN amorce codé par le même plasmide,

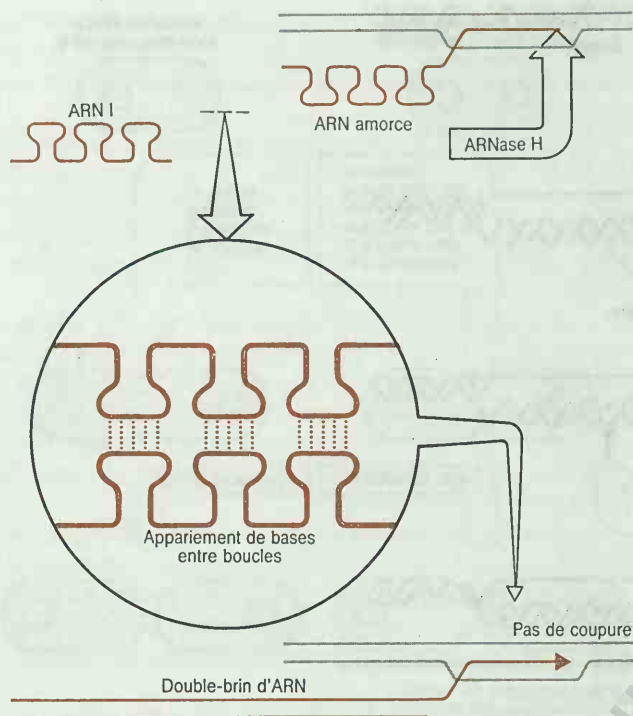


Figure 13.16

L'appariement avec l'ARN I change probablement la structure secondaire de la séquence de l'ARN amorce, empêchant le clivage et donc l'apparition d'une extrémité 3'-OH.

mais il lui est impossible de s'apparier avec l'ARN amorce codé par l'autre plasmide.

L'incompatibilité entre les plasmides du groupe de ColE1 est une conséquence du mécanisme utilisé pour initier la réplication. L'incompatibilité peut aussi être due à des mécanismes impliqués dans le partage des plasmides entre les cellules filles lors de la division bactérienne, mais peu de choses sont connues dans ce domaine. Il semble, en effet, que des mécanismes variés soient responsables du partage des plasmides entre les cellules filles.

Dans le plasmide pSC101, une région de 370pb, appelée *par*, est nécessaire à la ségrégation. La région *par* est un locus agissant en *cis* et qui ne code pas pour une protéine. Plus précisément, ce sont trois courts segments dans cette région qui assurent la partition. Les trois segments sont disposés de telle sorte que la séquence du segment central (si l'on considère les deux brins d'ADN séparément) peut s'apparier avec la séquence de l'un ou l'autre des segments flanquants, pour former une structure en épingle à cheveux. La délétion de l'un ou l'autre des trois segments empêche la partition.

Le phage P1, quand il est sous forme de prophage, est un plasmide dont la fonction *par* code pour une protéine qui reconnaît un site agissant en *cis*, appelé *incB* ; les mutations affectant l'incompatibilité entre plasmides sont localisées dans le locus *incB*.

Les sites, agissant en *cis*, impliqués dans la partition, sont peut-être les équivalents des centromères qui, chez les eucaryotes, permettent aux chromosomes de s'atta-

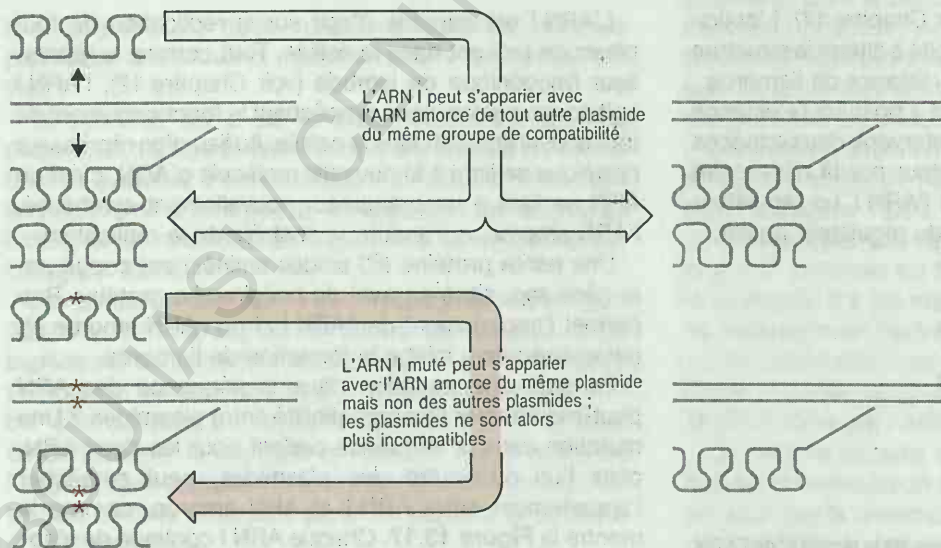


Figure 13.17

Des mutations dans la région codante pour l'ARN I et l'ARN amorce n'affectent pas leur capacité à s'apparier ; mais elles peuvent empêcher l'appariement avec un ARN complémentaire codé par un plasmide différent.

cher aux structures cellulaires. La compétition pour de tels sites pourrait être impliquée dans l'incompatibilité entre plasmides. Les protéines sont peut-être le maillon manquant entre partition et division cellulaire.

POUR EN SAVOIR PLUS

Des informations générales concernant les réplicons bactériens et eucaryotes sont apportées dans *Gene Expression* de Lewin,

[pp. 550-576 du volume 1, *Bacterial genomes*, et pp. 550-582 du volume 2, *Eucaryotic Chromosomes* (Wiley, New York, 1974 & 1980)]. Une origine de réplication fut isolée par Zyskind & Smith (*Proc. Nat. Acad. Sci. USA* 77, 2460-2464, 1980). Laskey & Harland (*Cell* 24, 283-284, 1981) émirent l'hypothèse de l'existence d'origines de réplication chez les eucaryotes. La réplication de l'ADN mitochondrial fut étudiée par Clayton (*Cell* 28, 693-705, 1982). Le contrôle de la régulation plasmidique fut revu par Scott (*Microbiol. Rev.* 48, 1-23, 1984). Le système de réplication de ColE1 fut élucidé par Tomizawa & Itoh (*Proc. Nat. Acad. Sci. USA* 78, 6096-6100, 1981).

CHAPITRE 14

L'APPAREIL DE RÉPLICATION DE L'ADN

La réplication de l'ADN bicaténaire est le résultat d'un mécanisme complexe qui implique de nombreuses activités enzymatiques. Tout d'abord, les brins parentaux sont séparés et maintenus (temporairement) à l'état simple brin. Puis, il y a initiation de la synthèse des molécules filles, molécules qui vont ensuite subir une élongation. Quand la réplication atteint l'extrémité terminale du réplicon, il se produit des réactions de jonction et de terminaison. Chacune de ces activités enzymatiques tient compte, pour assurer sa fonction, des contraintes imposées par la topologie complexe de l'ADN.

Les protéines responsables de ces différentes activités enzymatiques ne fonctionnent pas indépendamment les unes des autres ; elles sont probablement groupées dans une structure multiprotéique qu'on appelle le **réplisome**. Le réplisome n'est pas visible en tant que structure cellulaire libre (comme le ribosome, par exemple) ; il est formé au début de la réplication. Le réplisome est, plus précisément, *un complexe protéique associé à la structure particulière qu'adopte l'ADN au niveau de la fourche de réplication*.

Deux approches permettent de définir les pièces

nécessaires à l'appareil de réplication procaryote. D'une part, l'étude des mutants défectueux pour la réplication permet d'identifier des gènes dont le produit est impliqué soit directement, soit indirectement dans la réplication. D'autre part, il est possible d'étudier la réplication *in vitro* grâce à un système protéique défini. Grâce à cette première technique, l'appareil de réplication a été défini en terme de composants nécessaires à la réplication *in vitro*, c'est-à-dire en terme de produits biochimiques purs. L'étude des mutations *in vivo* a permis d'identifier les gènes codant pour ces composants.

Les enzymes capables de synthétiser de nouvelles molécules d'ADN à partir d'une matrice sont appelées **ADN polymérases**. Qu'elles soient procaryotes ou eucaryotes, les cellules possèdent de nombreuses activités polymérasiques. Seule l'une des polymérases est responsable de la fonction de réplication ; les autres jouent des rôles secondaires dans la réplication et/ou participent à la réparation de l'ADN afin de remplacer les séquences endommagées (voir Chapitre 15).

Il est commode de considérer qu'à chaque activité de synthèse de l'ADN correspond une ADN polymérase ;

mais l'existence d'une enzyme « réplicase » en tant qu'entité est encore un point discutable (le même problème a été posé en ce qui concerne l'ARN polymérase). L'activité réplicase de l'ADN bactérien est isolée sous la forme d'agrégats contenant de nombreuses « sous-unités » ; on ne sait pas encore clairement quelles sous-unités doivent être considérées comme composant de la réplicase elle-même lesquelles sont des facteurs auxiliaires. L'activité de synthèse de l'ADN est seulement l'une des fonctions du réplisome ; et peut-être, à cause de notre aptitude à mieux comprendre cette étape que d'autres, avons-nous tort de la considérer comme une activité indépendante.

Aucune ADN polymérase n'est capable d'initier la synthèse d'une chaîne désoxyribonucléotidique. Toutes les enzymes ne peuvent allonger que des molécules préexistantes. Il est donc essentiel que l'appareil de réplication ait un dispositif permettant l'initiation de la synthèse de l'ADN. Généralement, il y a synthèse d'une courte séquence ribonucléotidique qui disparaît par la suite.

A l'heure actuelle, chez les eucaryotes, ont été isolées de nombreuses enzymes impliquées dans la réplication. Il est très difficile de mettre au point des systèmes contenant tous les éléments nécessaires à la réplication, toujours est-il que la recherche d'un système, *in vitro*, permettant la réplication de molécules définies est en progrès. Jusqu'alors, seuls des systèmes nucléaires contenant tous les éléments nécessaires au mouvement de la fourche de réplication, ont été mis au point. L'initiation de la synthèse d'ADN à partir des origines de nouveaux réplicons n'a pas encore été obtenue *in vitro*.

LES ADN POLYMÉRASES, CES ENZYMES QUI FABRIQUENT L'ADN

Chez les eucaryotes, trois classes d'ADN polymérases ont été identifiées. Leurs propriétés sont récapitulées dans le **Tableau 14.1**. Est-ce que l'une d'entre elles peut être identifiée comme étant la réplicase ?

L'ADN polymérase alpha répond de la même façon *in vivo* et *in vitro* à l'adjonction d'inhibiteurs tels que la didésoxythymidine et l'aphidicoline. C'est la seule enzyme dont la quantité augmente durant la phase S. Ces observations nous permettent de considérer l'ADN polymérase alpha comme étant la réplicase de l'ADN nucléaire. Dans certains systèmes, il se trouve que l'ADN polymérase alpha a une activité d'amorçage dépendant de l'ARN (c'est-à-dire qu'elle permet la synthèse d'une amorce ARN qui sera allongée par l'ADN polymérase) probablement parce qu'une enzyme d'amorçage est étroitement associée à l'ADN polymérase.

Trois ADN polymérases ont été isolées chez *E. coli*. La première à avoir été découverte est l'ADN polymérase I. Elle joue un rôle, principalement, dans la réparation de l'ADN endommagé et, secondairement, dans la réplication semi-conservative. Rien n'est connu de la fonction *in vivo* de la seconde enzyme, l'ADN polymérase II. Quant à la troisième enzyme, l'ADN polymérase III, une protéine multimérique, elle est la réplicase responsable de la synthèse, *de novo*, de nouvelles molécules d'ADN. Les activités enzymatiques de ces trois enzymes sont résumées dans le **Tableau 14.2**.

Toutes les ADN polymérases, qu'elles soient procaryotes ou eucaryotes, ont le même type de fonction. Chacune

Tableau 14.1
Il existe trois ADN polymérases chez les mammifères.

Notez que les ADN polymérases des eucaryotes ne possèdent que l'activité de synthèse d'ADN, contrairement aux enzymes bactériennes qui possèdent de nombreuses autres activités, comme par exemple l'activité exonucléolytique (voir **Tableau 14.2**).

	ADN polymérase α	ADN polymérase β	ADN polymérase γ
Localisation Fonction	nucléaire réplication	nucléaire réparation	nucléaire et mitochondriale réplication mitochondriale
Activité relative Masse (daltons) N° de sous-unités	~ 80 % 110-220 000 plusieurs	10-15 % 45 000 1	2-15 % 60 000 1
Didésoxythymidine Aphidicoline	sans effet inhibiteur	inhibiteur sans effet	inhibiteur sans effet

Tableau 14.2

E. coli possède trois enzymes ADN polymérase.

	ADN polymérase I	ADN polymérase II	ADN polymérase III
Structure			
Daltons	109 000	120 000	supérieur à 250 000
Constitution	monomérique	inconnue	hétéromultimérique
Nombre/cellule	400	inconnue	10-20
Activités enzymatiques			
Élongation 5'-3'	oui	oui	oui
Exonucléase 3'-5'	oui	oui	oui
Exonucléase 5'-3'	oui	non	oui
Mutants			
Loci mutés	<i>polA</i>	<i>polB</i>	<i>polC, dnaN, dnaX, dnaZ, dnaQ</i>
Phénotype mutant	défectueux pour la réplication	aucun	empêche la réplication
Létalité	la viabilité est réduite uniquement si l'activité exonucléasique est affectée	sans effet	léthal conditionnel

a la capacité d'allonger une chaîne d'ADN en ajoutant, un par un, des nucléotides à l'extrémité 3'-OH, comme l'illustre la **Figure 14.1**. Le choix du nucléotide, qui est ajouté à la chaîne en cours d'élongation, est dicté par l'appariement de celui-ci avec la molécule qui sert de matrice.

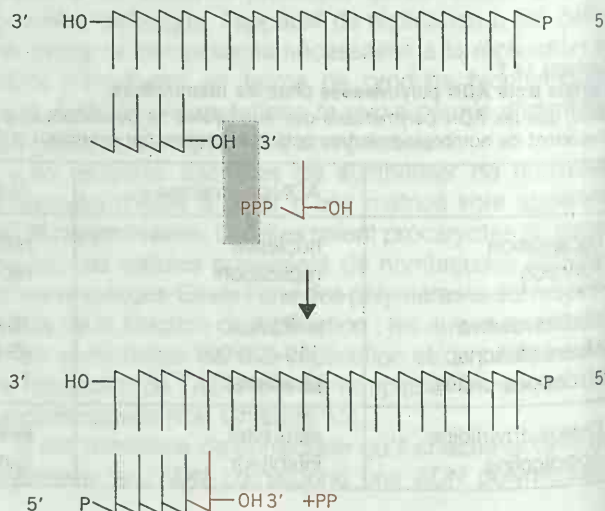
La fidélité de la réplication pose les mêmes problèmes que ceux rencontrés pour expliquer (par exemple) la justesse et l'exactitude de la traduction. Elle repose sur la spécificité de l'appariement. Si l'on considère les interactions impliquées dans l'appariement, l'on s'attend à une fréquence d'erreurs de 10^{-4} à 10^{-5} par paire de bases répliquée. La fréquence réelle d'erreurs chez les bactéries est de 10^{-8} à 10^{-10} , ce qui correspond à environ une erreur pour 1 000 cycles de réplication ou, en d'autres termes, à un taux d'erreurs d'environ 10^{-6} par gène et par génération.

L'enzyme doit tester la spécificité de complémentarité entre les bases avant ou après l'appariement (ou les deux). Il est possible qu'elle examine la base qui arrive quant à sa complémentarité avec la base à laquelle elle va s'apparier en reconnaissant, par exemple, des interactions chimiques précises. Ce serait une étape de contrôle « *présynthétique* ». Ou alors, l'enzyme ne peut vérifier le bon appariement qu'après coup, et dans ce cas, si une erreur a été commise, l'enzyme la corrigera en détachant la dernière base arrivée sur la molécule en cours d'élongation.

Toutes les enzymes bactériennes possèdent une activité 3'-5' exonucléasique, qui agit donc dans le sens

Figure 14.1

La synthèse d'ADN se fait à partir de l'extrémité 3'-OH de la chaîne naissante dans la direction 5'-3'. Le précurseur de la synthèse d'ADN est un nucléoside triphosphate qui, lors de la réaction, perd ses deux groupements phosphate terminaux.



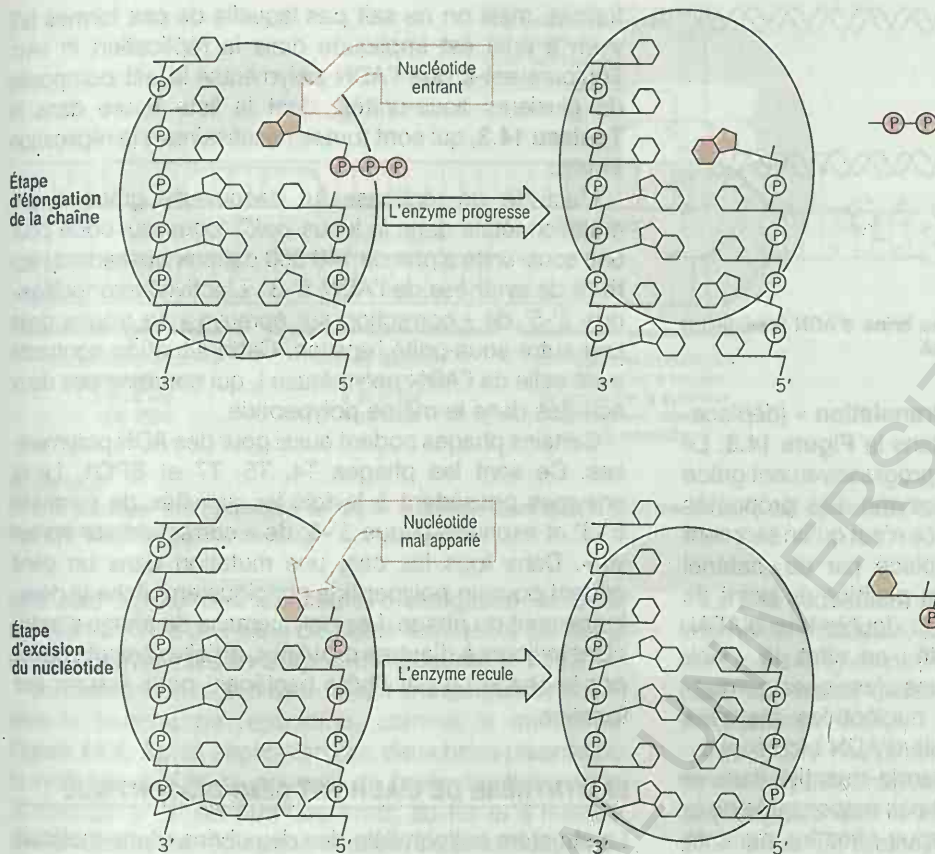


Figure 14.2

Les ADN polymérases bactériennes vérifient l'appariement des bases à la fin de la chaîne en cours d'élongation et excisent le dernier nucléotide ajouté s'il est mal apparié.

opposé à la synthèse de l'ADN, et qui semble exercer une fonction post-synthétique de « **correction sur épreuve** » (proofreading) du brin néosynthétisé. Les enzymes eucaryotes ne possédant pas cette activité, il doit exister un autre mécanisme permettant de contrôler le taux d'erreurs dans la réplication.

La fonction de « correction sur épreuve » est illustrée dans la Figure 14.2. L'élongation de la molécule d'ADN se fait de la façon suivante : un nucléotide précurseur se positionne près de l'extrémité 3'-OH de la chaîne naissante à laquelle il se lie, l'enzyme va alors se déplacer d'une paire de bases, prête à recevoir le prochain nucléotide qui se présentera. Si une erreur a été commise, l'enzyme revient en arrière, excise la dernière base ajoutée et libère la place pour accueillir un autre nucléotide précurseur.

L'ADN polymérase I est une enzyme douée de multiples

activités. C'est un polypeptide de 109 000 daltons, codé par le locus *polA*. Après traitement protéolytique, la protéine est clivée en deux fragments. Le plus grand fragment (76 000 daltons) contient les activités polymérasique 5'-3' et exonucléasique 3'-5'. Le petit fragment (36 000 daltons) possède l'activité exonucléasique 5'-3', activité qui est indépendante des fonctions de synthèse et de « correction sur épreuve ». Cette activité exonucléasique 5'-3' permet l'excision de petits groupes de nucléotides, allant jusqu'à environ dix nucléotides.

L'ADN polymérase I est la seule enzyme capable d'initier la réplication *in vitro* à partir d'une coupure sur la molécule d'ADN. L'enzyme allonge l'extrémité 3'-OH, là où la liaison phosphodiester a été cassée sur une molécule d'ADN bicaténaire. La synthèse d'un nouveau brin d'ADN déplace progressivement le brin homologue préexistant.

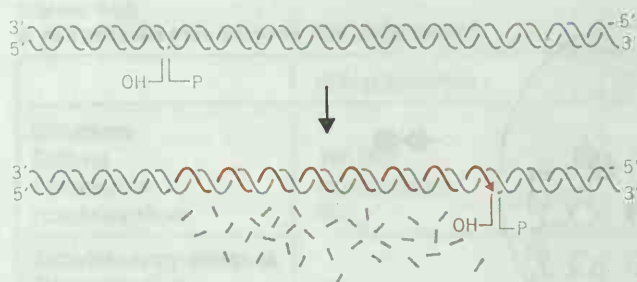


Figure 14.3

La « nick translation » remplace un des brins d'ADN préexistant par du matériel nouvellement synthétisé.

Ce procédé est appelé « **nick translation** » (déplacement de coupure), il est illustré dans la **Figure 14.3**. Le brin qui est remplacé est dégradé progressivement grâce à l'activité exonucléasique de l'enzyme. Les propriétés de l'ADN ne sont pas modifiées, si ce n'est qu'un segment d'un des deux brins a été remplacé par du matériel nouvellement synthétisé, et que la position du point de cassure a été déplacée le long du double-brin d'ADN. Ceci est d'une très grande utilité ; en effet, la « nick translation » est une des techniques classiques permettant l'introduction, *in vitro*, de nucléotides marqués radioactivement dans une molécule d'ADN bicaténaire.

Bien qu'elle soit trouvée en grande quantité dans la bactérie, l'ADN polymérase I n'est pas responsable de la réplication qui n'est pas interrompue chez les mutants *polA* déficients en cette activité. Les activités de synthèse 5'-3' et exonucléasique 3'-5' sont probablement utilisées, *in vivo*, pour combler de courtes régions monocaténares dans une molécule d'ADN bicaténaire. De telles régions apparaissent lors de la réplication, et aussi, lorsque certaines bases ont été endommagées puis excisées de la molécule (voir Chapitre 15).

L'ADN polymérase III a été isolée sous différentes

formes, mais on ne sait pas laquelle de ces formes (s'il y en a une) est impliquée dans la réplication *in vivo*. Toujours est-il que l'ADN polymérase III est composée de plusieurs sous-unités, dont la liste figure dans le **Tableau 14.3**, qui sont toutes nécessaires à la réplication *in vitro*.

L'activité de réplicase fut découverte grâce à une mutation létale dans le locus *polC*, locus qui code pour une sous-unité alpha de 140 000 daltons possédant l'activité de synthèse de l'ADN 5'-3'. L'activité exonucléasique 3'-5' de « correction sur épreuve » se trouve dans une autre sous-unité, epsilon. Cette situation contraste avec celle de l'ADN polymérase I, qui combine ces deux activités dans le même polypeptide.

Certains phages codent aussi pour des ADN polymérases. Ce sont les phages T4, T5, T7 et SPO1. Leurs enzymes possèdent à la fois les activités de synthèse 5'-3' et exonucléasique 3'-5' de « correction sur épreuve ». Dans tous les cas, une mutation dans un gène codant pour un polypeptide phagique empêche le développement du phage. Les polymérases de phage s'associent toujours à d'autres protéines, qu'elles soient codées par le phage ou par l'hôte bactérien, pour assurer leur fonction.

LA SYNTHÈSE DE L'ADN EST SEMI-DISCONTINUE

La structure antiparallèle des deux brins d'une molécule d'ADN bicaténaire pose un problème pour la réplication. Au fur et à mesure que la fourche de réplication avance, les deux molécules filles doivent être synthétisées à partir des deux brins parentaux. La fourche de réplication se déplace dans la direction 5'-3' sur l'un des brins et dans la direction 3'-5' sur l'autre brin, alors que la synthèse de l'ADN, elle, s'effectue toujours d'une extrémité 5' vers une extrémité 3'. Ce problème est résolu grâce à la

Tableau 14.3

L'ADN polymérase II a plusieurs sous-unités potentielles.

L'« holoenzyme » est constituée de toutes les sous-unités ; les autres formes de l'enzyme (cœur pol III, pol III', pol III'') sont le résultat de différentes combinaisons de sous-unités.

Sous-unité	Daltons	Autres noms et/ou gène	Fonction enzymatique
α	140 000	<i>polC</i> (<i>dnaE</i>)	Synthèse d'ADN
β	40 000	<i>dnaN</i> = facteur I = coPol III*	} Facteur II
δ	32 000	<i>dnaX</i> = facteur III	
γ	52 000	<i>dnaZ</i>	
ϵ	25 000	<i>dnaQ</i>	
θ	10 000		
τ	83 000		exonucléase 3'-5'

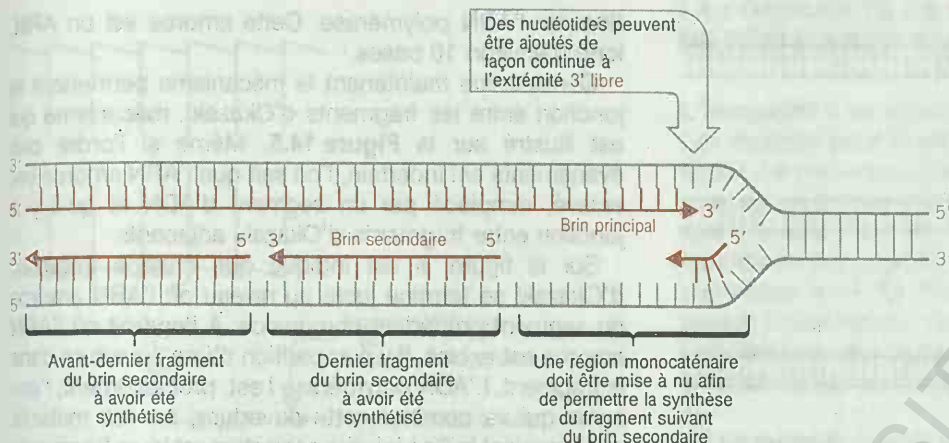


Figure 14.4

Le brin principal est synthétisé de façon continue alors que le brin secondaire est synthétisé de façon discontinue.

synthèse, sur le brin en cours d'élongation dans la direction 3'-5', d'une série de petits fragments qui sont, en réalité, synthétisés dans la polarité habituelle, 5' vers 3'.

Considérons la région se trouvant immédiatement derrière la fourche de réplication, comme le montre la Figure 14.4. Après séparation des deux brins parentaux, la synthèse d'ADN se poursuit de façon continue dans la direction 5'-3' sur l'un des brins, au fur et à mesure que la fourche de réplication avance. Le brin néosynthétisé est appelé **brin principal** (leading strand). En revanche, l'autre brin parental reste à découvert le temps qu'il y ait une zone suffisamment longue pour permettre la synthèse d'un segment dans la direction opposée au mouvement de la fourche, c'est-à-dire dans la direction 5'-3'. De petits fragments sont ainsi synthétisés à la suite, chacun dans la direction 5'-3' ; puis, ils sont réunis entre eux afin de créer un brin intact. Ce brin est appelé **brin secondaire** (lagging strand) et le mécanisme qui l'engendre est appelé **réplication discontinue**.

La réplication discontinue peut être étudiée grâce à un bref marquage radioactif. L'ADN nouvellement synthétisé sur le brin secondaire est marqué sous forme de petits fragments qui sédimentent à 7-11S, ce qui correspond à des séquences d'environ 1 000-2 000 bases de long. Ces fragments sont appelés **fragments d'Okazaki**. Ils sont trouvés dans tout ADN en cours de réplication, aussi bien chez les procaryotes que chez les eucaryotes. Après de longues périodes d'incubation, le marquage radioactif est présent sur de plus longs segments d'ADN. Ces longs segments représentent des fragments d'Okazaki reliés entre eux par liaison covalente.

Le brin secondaire *doit* être synthétisé sous forme de fragments d'Okazaki. Le mécanisme de synthèse du brin principal a longtemps été discuté, à savoir si la réplication était, elle aussi, discontinue sur ce brin, ou bien tout simplement continue.

Tout ADN nouvellement synthétisé est trouvé sous forme de petits fragments chez *E. coli*. À première vue, cela suggère que la réplication est discontinue sur les deux brins. En réalité, tous ces fragments ne sont pas des fragments d'Okazaki ; certains sont des pseudo-fragments, engendrés par des cassures dans la molécule d'ADN synthétisée en fait de façon continue.

Ces cassures sont le résultat de l'incorporation d'uracile dans l'ADN, à la place de la thymine. Quand les uraciles sont excisés de la molécule grâce au système de réparation, le brin principal est sous forme de fragments jusqu'à ce que des thymines soient insérées aux points de cassures.

Chez *E. coli*, le brin principal est synthétisé de façon continue alors que le brin secondaire est synthétisé de façon discontinue. Cela est sûrement valable pour d'autres systèmes. Ce mode de synthèse d'ADN est appelé **réplication semi-discontinue**.

LES FRAGMENTS D'OKAZAKI SONT AMORCÉS PAR DES ARN

Nous avons déjà mentionné que toutes les ADN polymérases connues, procaryotes comme eucaryotes, allongent une chaîne désoxyribonucléique à partir d'une extré-

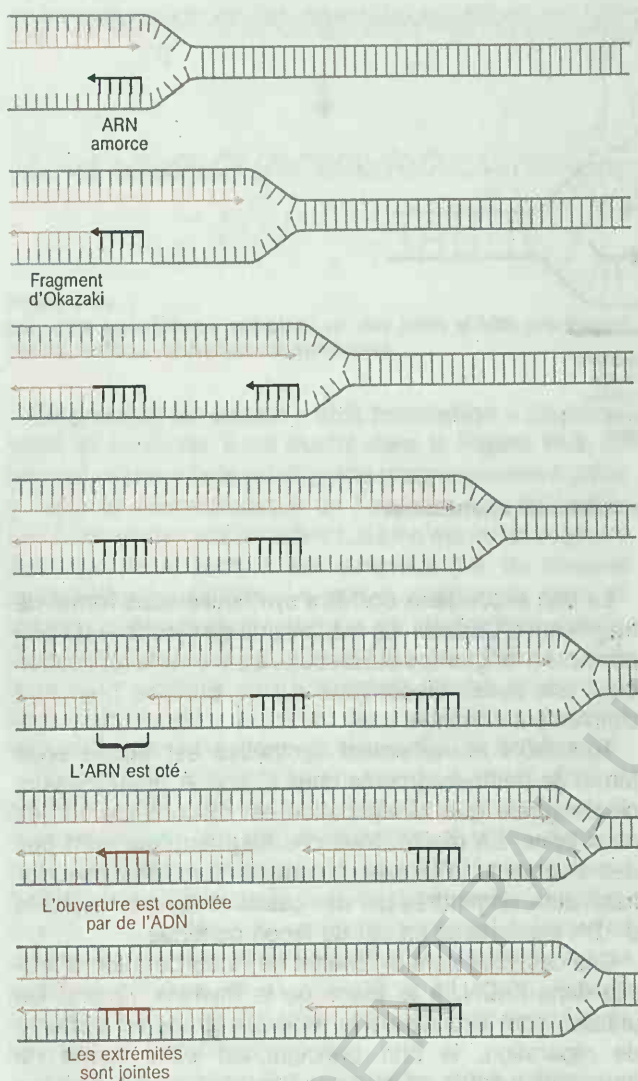


Figure 14.5

Les différentes étapes permettant la liaison des fragments d'Okazaki sont la suppression de l'ARN, le comblement de l'ouverture et la jonction des extrémités.

mité 3'-OH libre, mais qu'aucune d'entre elles ne peut initier la synthèse. Il doit donc se produire de nombreux événements d'initiation étant donné que chaque fragment d'Okazaki, pour être synthétisé, doit tout d'abord être initié. Chaque fragment d'Okazaki est initié grâce à une amorce, c'est-à-dire grâce à une courte séquence qui possède une extrémité 3'-OH libre permettant l'élonga-

tion par l'ADN polymérase. Cette amorce est un ARN, long d'environ 10 bases.

Considérons maintenant le mécanisme permettant la jonction entre les fragments d'Okazaki, mécanisme qui est illustré sur la **Figure 14.5**. Même si l'ordre des événements est incertain, l'on sait que l'ARN amorce est enlevé, remplacé par un segment d'ADN et qu'il y a jonction entre fragments d'Okazaki adjacents.

Sur la figure, il est indiqué que chaque fragment d'Okazaki se termine juste au niveau où l'ARN amorce du segment précédent commence. A l'endroit où l'ARN amorce est excisé, il y a apparition d'une ouverture dans le fragment. L'ADN polymérase I est, probablement, l'enzyme qui va combler cette ouverture, car les mutants *polA* peuvent faillir à joindre correctement leurs fragments d'Okazaki. En fait, cette enzyme doit utiliser simultanément son activité exonucléasique 5'-3', afin d'exciser l'ARN amorce, et son activité polymérasique, afin d'allonger la molécule d'ADN à partir de l'extrémité 3'-OH du fragment d'Okazaki suivant. Ce mécanisme est l'équivalent de la « nick translation », mis à part que l'ADN nouvellement synthétisé remplace un segment d'ARN au lieu d'un segment d'ADN.

Une fois l'ARN excisé et remplacé, les fragments d'Okazaki adjacents doivent être reliés entre eux. L'extrémité 3'-OH de chaque fragment est adjacente à l'extrémité 5'-phosphate du fragment précédent. L'enzyme permettant la liaison entre les fragments est l'**ADN ligase** (dont le rôle de jonction a déjà été mentionné lors de la construction de molécules d'ADN *in vitro*). Les ligases sont présentes aussi bien chez les procaryotes que chez les eucaryotes. Les mutants *lig-* sont incapables de relier les fragments d'Okazaki entre eux.

Les ligases de T4 et d'*E. coli* ont la propriété de relier les extrémités 3'-OH et 5'-phosphate entre elles, comme l'illustre la **Figure 14.6**. La réaction se fait en deux étapes, impliquant un complexe enzyme-AMP. L'AMP du complexe enzymatique se lie à l'extrémité 5'-phosphate, et permet la formation d'une liaison phosphodiester avec l'extrémité 3'-OH, ce qui a pour effet de libérer à la fois l'enzyme et l'AMP.

La découverte de ces événements soulève de nombreuses questions quant à la spécificité de la synthèse des fragments d'Okazaki. Quelle est l'enzyme responsable de la synthèse des ARN amorces ? Son action commence-t-elle et finit-elle à des séquences précises, ou est-elle déterminée par l'organisation de la fourche de réplication ? La synthèse d'un fragment d'Okazaki continue-t-elle jusqu'à ce qu'elle rencontre le fragment adjacent, ou existe-t-il des sites de terminaison ?

LA COMPLEXITÉ DE L'APPAREIL DE RÉPLICATION BACTÉRIEN

L'incapacité à se répliquer est fatale pour toute cellule. Les mutants pour la répllication sont donc des mutants létaux conditionnels, c'est-à-dire qu'ils peuvent se répliquer sous conditions permissives (incubation à température normale), mais sont incapables de répllication sous conditions non permissives (température d'incubation supérieure à 42 °C). Ces mutants thermosensibles ont permis l'identification d'une série de loci, les gènes *dna*. Les mutants *dna* sont divisés en deux catégories en fonction de leur comportement vis-à-vis de la température.

Les mutants à arrêt lent sont capables de terminer le cycle de répllication en cours, mais pas d'en commencer un autre. Ils sont défectueux pour l'initiation de la répllication à l'origine. Peu de choses sont connues en ce qui les concerne (voir Tableau 14.6), mais il est certain que plusieurs protéines différentes sont nécessaires pour initier le cycle de répllication.

Les mutants à arrêt rapide cessent la répllication dès que la température est critique. Ces mutants sont défectueux pour l'un des éléments composant le replisome (voir Tableau 14.4). Ils pourraient également être défectueux pour l'initiation de la répllication, bien qu'alors, il soit difficile d'expliquer leur phénotype.

Certains mutants à arrêt rapide sont incapables de produire les précurseurs nécessaires à la synthèse de l'ADN. Ce sont, par exemple, les mutants pour le gène *nrdA*, initialement appelé *dnaF* ; ce gène code pour une sous-unité de la ribonucléotide réductase, qui est essentielle pour la formation des précurseurs. Cela montre l'importance d'un bon approvisionnement en précurseurs pour la synthèse de l'ADN. En d'autres termes, un mutant portant le phénotype Dna n'est pas forcément un mutant dont le produit du gène muté est directement impliqué dans la répllication.

Bien que des efforts aient été fournis pour isoler un grand nombre de mutants *dna*, les mutations découvertes à l'heure actuelle ne permettent pas de couvrir toutes les fonctions nécessaires à la répllication. Des protéines supplémentaires, codées par des gènes inconnus, sont requises dans les systèmes de répllication *in vitro* (voir Tableau 14.5).

Lorsque l'on teste des extraits d'*E. coli* quant à leur capacité à synthétiser de l'ADN, l'activité prédominante est celle de l'ADN polymérase I. En fait, l'activité de cette enzyme est si importante qu'il est impossible de détecter les activités des enzymes responsables de la répllication

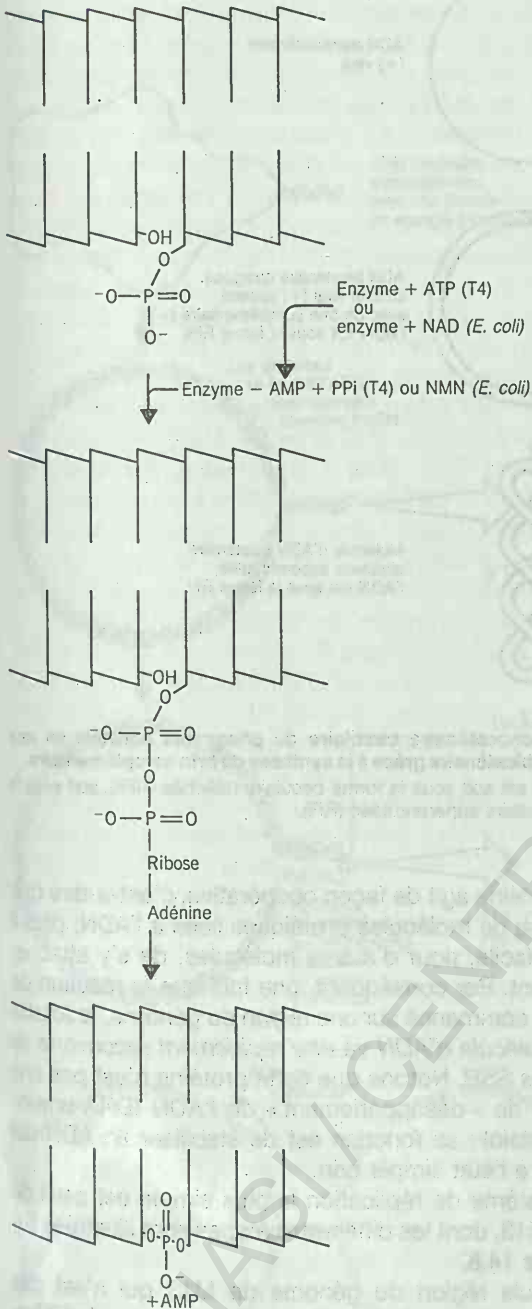


Figure 14.6
L'ADN ligase lie deux nucléotides adjacents en utilisant un intermédiaire enzyme-AMP.

Les enzymes d'*E. coli* et de T4 utilisent différents cofacteurs. L'enzyme d'*E. coli* utilise le NAD (dinucléotide adénine nicotinamide) comme cofacteur alors que T4 utilise l'ATP.

de l'ADN. Afin de développer, *in vitro*, des systèmes permettant de suivre la réplication, les extraits cellulaires sont préparés à partir de mutants *polA*.

Afin d'identifier les composants nécessaires à la réplication *in vitro*, l'on doit s'assurer que la synthèse d'ADN est bien le reflet du processus de réplication *in vivo*, et non celui d'une autre activité, comme par exemple la réparation de l'ADN. Différents modèles peuvent être utilisés pour établir des systèmes *in vitro*, reflétant soit la synthèse des brins principaux, soit la synthèse des brins secondaires.

Le rôle de chaque composant peut être exploré de deux façons. D'une part, le système peut être divisé en différentes fractions qui seront purifiées puis reconstituées. D'autre part, le système de réplication peut être préparé à partir des mutants *dna* puis rendu opérationnel alors que le produit du gène muté est inactif. En effet, des extraits provenant de cellules sauvages seront testés quant à leur capacité à restaurer l'activité manquante. C'est ce que l'on appelle un **test de complémentation *in vitro***, il peut être utilisé pour purifier la protéine codée par le locus *dna*.

INITIATION DE LA SYNTHÈSE D'UNE MOLÉCULE D'ADN MONOCATÉNAIRE

La production d'un fragment d'Okazaki est le résultat de la synthèse d'une molécule d'ADN, complémentaire de la région simple brin mise à nu grâce au mouvement de la fourche de réplication. Trois phages, dont la molécule d'ADN est monocaténaire, ont été utilisés pour étudier les mécanismes d'initiation et d'élongation du brin secondaire.

Le génome de chacun de ces phages consiste en une molécule d'ADN monocaténaire circulaire, appelée **brin viral** ou brin (+). La **Figure 14.7** nous montre que la synthèse du brin **complémentaire** ou brin (-) transforme le génome en une molécule d'ADN circulaire bicaténaire, qui est alors sous **forme répllicative (RF)**.

Même si le génome de tels phages montre très peu d'appariement en son sein et par conséquent reste sous forme monocaténaire, la molécule d'ADN qui le définit n'est pas, en elle-même, suffisante pour permettre la réplication. En effet, l'ADN doit, en premier lieu, être recouvert d'une protéine de l'hôte se liant spécifiquement à l'ADN monocaténaire, la protéine **SSB (single-strand binding protein)** : protéine se liant à l'ADN simple brin). Cette protéine est un tétramère de 74 000 daltons qui se lie de façon coopérative à l'ADN simple brin, et le maintient à l'état étendu.

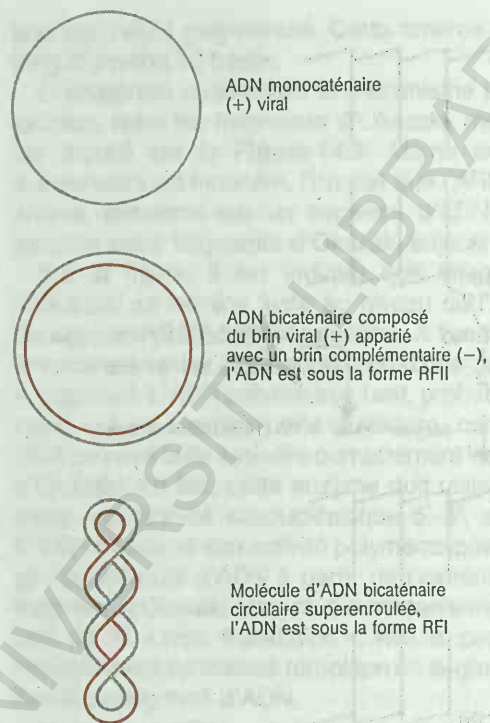


Figure 14.7

L'ADN monocaténaire circulaire du phage est converti en une molécule bicaténaire grâce à la synthèse du brin complémentaire. L'ADN RF est soit sous la forme circulaire relâchée (RFII), soit sous la forme circulaire superenroulée (RFI).

La protéine agit de façon coopérative, c'est-à-dire que plus il y a de molécules protéiques liées à l'ADN, plus il devient facile, pour d'autres molécules, de s'y attacher également. Par conséquent, une fois que la réaction de liaison a commencé sur une région du génome, *la totalité de la molécule d'ADN va être rapidement recouverte de protéines SSB*. Notons que cette protéine n'est pas une protéine de « désappariement » de l'ADN (DNA-unwinding protein) ; sa fonction est de stabiliser un ADN qui est déjà à l'état simple brin.

Le système de réplication le plus simple est celui du phage M13, dont les différentes étapes sont illustrées sur la **Figure 14.8**.

La seule région du génome de M13 qui n'est pas recouverte de protéines SSB est une séquence de 59 bases qui forme une boucle en épingle à cheveux. Cette région bicaténaire est un signal d'initiation. L'ARN polymérase de l'hôte initie la synthèse de l'ARN six bases avant le début de la boucle en épingle à cheveux. On ne sait pas encore comment l'ARN polymérase reconnaît la boucle en épingle à cheveux en tant que site d'initiation ;

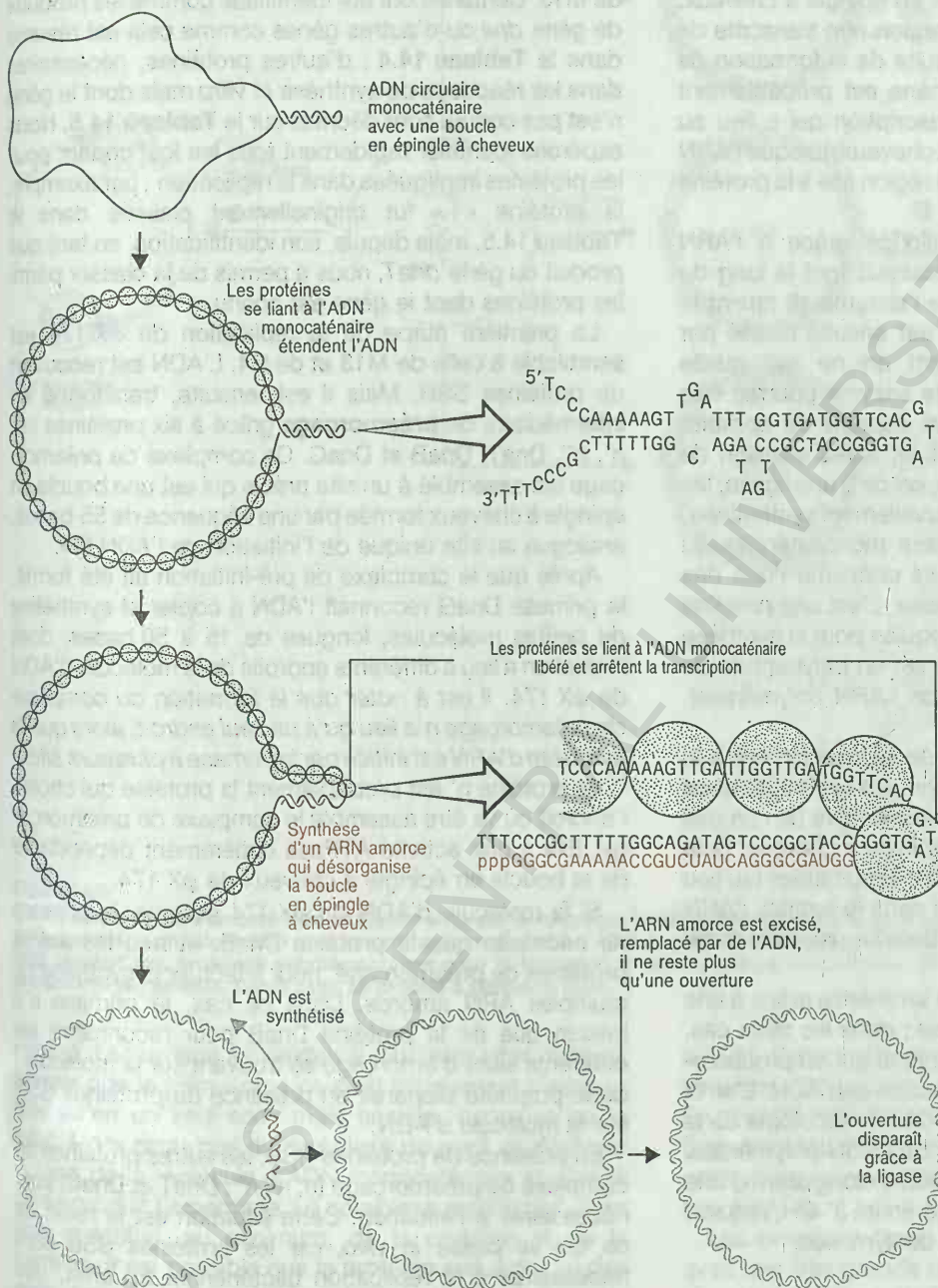


Figure 14.8

La synthèse du brin complémentaire de M13 est possible grâce à une matrice d'ADN viral recouverte de protéines SSB. Les différentes étapes de cette synthèse sont l'initiation grâce à un ARN amorce, l'élongation de cette amorce par de l'ADN, l'excision de l'amorce, son remplacement par de l'ADN et enfin la liaison des deux extrémités libres.

il est possible que d'autres protéines participent à cette reconnaissance.

Un brin d'ARN de 20-30 résidus est synthétisé, ce qui a pour effet de détruire la boucle en épingle à cheveux. Les protéines SSB se lient à la région non transcrite de l'épingle à cheveux libérée par suite de la formation de l'hybride ADN-ARN. Ce phénomène est probablement responsable de l'arrêt de la transcription qui a lieu au sommet de la boucle en épingle à cheveux, lorsque l'ARN polymérase vient au contact de la région liée à la protéine SSB (Figure 14.8).

L'ARN amorce est ensuite allongé grâce à l'ARN polymérase III. La synthèse se poursuit tout le long du génome circulaire jusqu'à ce que l'enzyme ait atteint le début de l'ARN amorce. L'ARN est ensuite excisé par une action exonucléolytique dont on ne sait quelle enzyme en est responsable. Cette enzyme pourrait être l'ADN polymérase I, qui est aussi capable de combler l'ouverture laissée dans la molécule, après excision de l'ARN. Il reste simplement à relier, grâce à une ligase, les deux extrémités de la molécule nouvellement synthétisée.

La réplication de l'ADN circulaire monocaténaire du phage G4 est différente en ce qui concerne l'une des étapes. Au lieu de l'ARN polymérase, c'est une protéine codée par le gène *dnaG* qui est requise pour la synthèse d'un ARN amorce. Cette enzyme est un polypeptide de 60 000 daltons (bien plus petit que l'ARN polymérase). C'est une **primase**.

Une molécule protéique DnaG se lie directement à la boucle en épingle à cheveux. La primase synthétise alors un ARN de 29 bases qui est complémentaire de l'un des deux brins formant la boucle en épingle à cheveux. Si la synthèse de l'ADN et l'initiation sont simultanées (au lieu d'être deux étapes bien séparées dans le temps), l'ARN amorce est plus court, pouvant atteindre une taille aussi petite que six bases.

Un ARN amorce peut donc être synthétisé grâce à une ARN polymérase ou à une primase ; dans les deux cas, la synthèse commence à un site unique qui est probablement défini par la structure secondaire de l'ADN. L'arrêt de la synthèse est probablement dû à la topologie de la région d'initiation et/ou à l'action de l'ADN polymérase. Quel que soit le phage considéré, l'élongation d'une molécule d'ADN, à partir de son extrémité 3'-OH, requiert toujours les mêmes mécanismes de synthèse.

LE MOUVEMENT DU PRIMOSOME

Un meilleur exemple pour étudier l'initiation des fragments d'Okazaki est peut-être illustré par le phage ϕ X 174, dont la synthèse d'ADN est initiée à plusieurs endroits du génome. Le mécanisme de réplication de

ϕ X 174 est comparé à ceux de M13 et de G4 sur la **Figure 14.9**. Beaucoup plus de protéines sont requises pour répliquer l'ADN de ϕ X 174 que pour celui de G4 et de M13. Certaines ont été identifiées comme les produits de gène *dna* ou d'autres gènes comme cela est résumé dans le **Tableau 14.4** ; d'autres protéines, nécessaires dans les réactions de synthèse *in vitro* mais dont le gène n'est pas connu, sont décrites sur le **Tableau 14.5**. Nous espérons identifier rapidement tous les loci codant pour les protéines impliquées dans la réplication ; par exemple, la protéine « i » fut originellement classée dans le **Tableau 14.5**, mais depuis, son identification, en tant que produit du gène *dnaT*, nous a permis de la classer parmi les protéines dont le gène est connu.

La première étape de la réplication de ϕ X 174 est semblable à celle de M13 et de G4. L'ADN est recouvert de protéines SSB. Mais il est, ensuite, transformé en intermédiaire de **préamorçage** grâce à six protéines : n, n', n'', DnaT, DnaB et DnaC. Ce complexe de préamorçage est assemblé à un site précis qui est une boucle en épingle à cheveux formée par une séquence de 55 bases, analogue au site unique de l'initiation de l'ADN-G4.

Après que le complexe de pré-initiation ait été formé, la primase DnaG reconnaît l'ADN à copier et synthétise de petites molécules, longues de 15 à 50 bases, dont l'initiation a lieu à différents endroits de la molécule d'ADN de ϕ X 174. *Il est à noter que la formation du complexe de préamorçage n'a lieu qu'à un seul endroit, alors que la synthèse d'ARN est initiée par la primase à plusieurs sites.*

La protéine n' est probablement la protéine qui choisit l'endroit où va être assemblé le complexe de préamorçage. Elle a une activité ATPase entièrement dépendante de la boucle en épingle à cheveux de ϕ X 174.

Si la molécule d'ADN de ϕ X 174 est nue, la primase ne nécessite que la protéine DnaB, et non les autres protéines de préamorçage, pour effectuer la synthèse de multiples ARN amorce. Dans ce cas, la primase n'a besoin que de la protéine DnaB pour reconnaître les différents sites d'amorçage se trouvant sur la molécule ; cette propriété disparaît en présence de protéines SSB sur la molécule d'ADN.

En présence de protéines SSB, les autres protéines du complexe de préamorçage (n, n', n'', DnaT et DnaC) sont nécessaires à l'initiation. Cette situation est le reflet de ce qui se passe *in vivo*, car les protéines SSB sont nécessaires à la réplication bactérienne. En effet, une mutation dans le gène codant pour la protéine SSB entraîne un phénotype de mutant à arrêt rapide défini précédemment.

Le complexe protéique impliqué dans la réaction d'amorçage est appelé le **primosome**. Le fait que le préamorçage n'ait lieu qu'à un seul site alors qu'ultérieu-

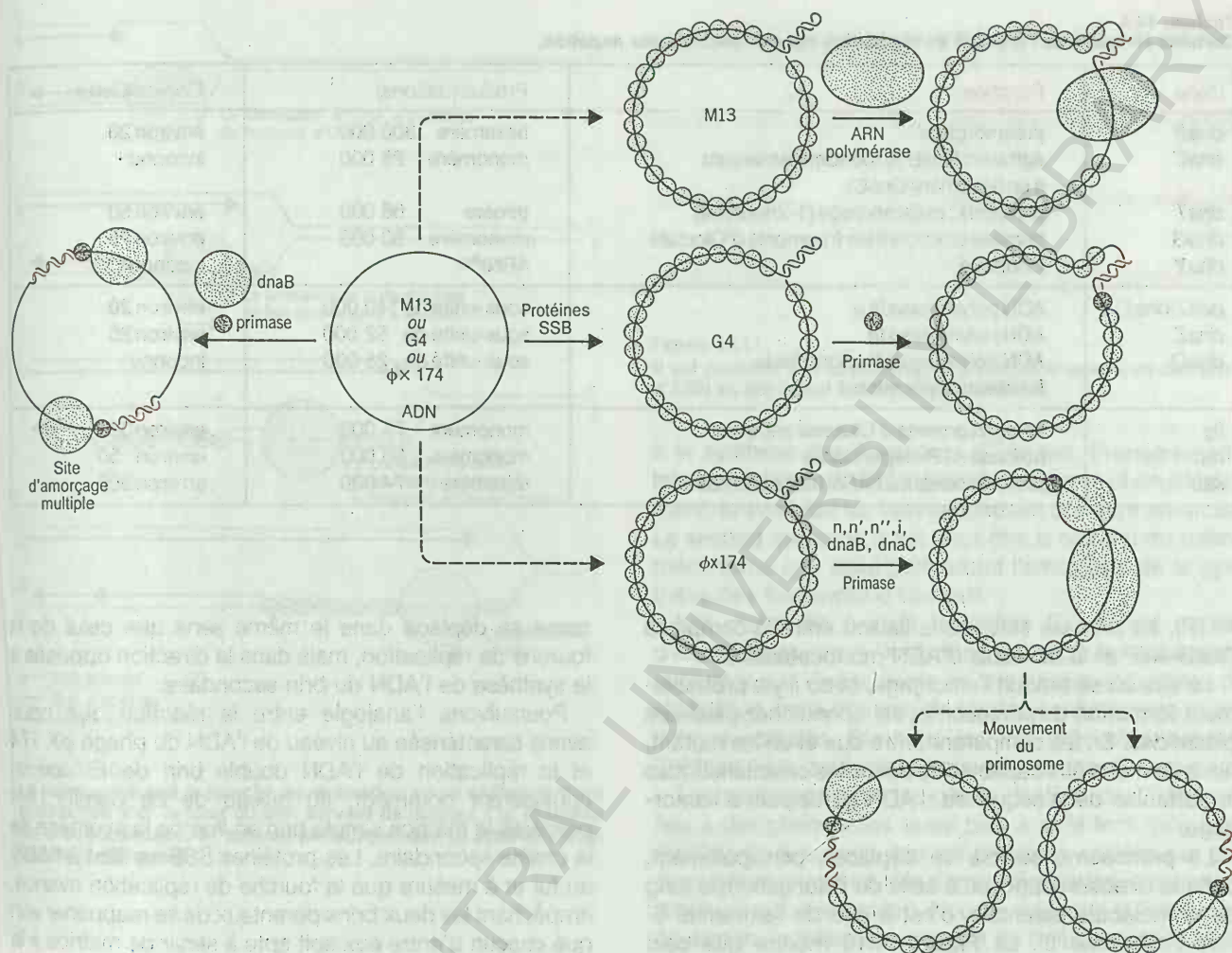


Figure 14.9

L'ADN libre peut être amorcé à de multiples sites par le complexe DnaB-primase. L'ADN de M13 ou G4 recouvert de protéines SSB n'est amorcé qu'à un seul site. L'ADN de ϕ X174 recouvert de protéines SSB requiert des protéines supplémentaires pour la formation à un site unique du complexe de pré-amorçage ; ensuite, l'amorçage se produit à plusieurs sites.

rement, l'amorçage se produit à plusieurs sites, incite à penser que le primosome ne peut initialement s'assembler qu'en un seul site ; mais ensuite, certaines ou la totalité des protéines du complexe peuvent se déplacer le long de la molécule d'ADN monocaténaire et aller là où se produit l'amorçage. Le complexe mobile doit donc être constitué, au minimum, de la protéine et de la primase. Il est possible que la protéine soit aussi un des composants du complexe mobile, car elle est capable de détacher les protéines SSB de la molécule d'ADN ; cette propriété semble essentielle pour qu'il y ait mouvement sur la molécule d'ADN.

Le comportement de la protéine DnaB est peut-être également important pour le déplacement sur la molécule

d'ADN du complexe protéique. Il semble y avoir un hexamère de protéines DnaB par fourche de réplication. En présence d'ATP, l'hexamère forme un complexe avec 6 sous-unités de protéine DnaC ; et c'est probablement cette association qui permet la liaison initiale des protéines à l'ADN.

Les protéines DnaB se lient à l'ADN monocaténaire avec une très grande affinité en présence d'ATP. Si l'on remplace l'ATP par de l'ADP, il se produit un changement allostérique qui réduit l'affinité des protéines DnaB pour l'ADN. L'hydrolyse d'ATP, après la réaction d'amorçage, doit être nécessaire pour permettre aux protéines DnaB de se dissocier de l'ADN afin de se déplacer jusqu'à un autre site d'amorçage (ou l'ADP est remplacé par de

Tableau 14.4
Certains éléments de l'appareil de réplication ont été identifiés par mutation.

Gène	Fonction	Produit (daltons)	Copies/Cellule
<i>dnaB</i> <i>dnaC</i>	préamorçage agit avec DnaB (6 monomères se lient à un hexamère DnaC)	hexamère 300 000 monomère 25 000	environ 20 inconnu
<i>dnaT</i> <i>dnaG</i> <i>dnaY</i>	protéine i : préamorçage (1-2/fourche) primase amorçant les fragments d'Okazaki inconnue	trimère 66 000 monomère 60 000 ARNt ^{Arg}	environ 50 environ 75 inconnu
<i>polC (dnaE)</i> <i>dnaZ</i> <i>dnaQ</i>	ADN polymérase III α ADN polymérase III ADN polymérase III ϵ contrôle la fidélité de réplication	sous-unité de 140 000 sous-unité de 52 000 sous-unité de 25 000	environ 20 environ 20 inconnu
<i>lig</i> <i>rep</i> <i>ssb</i>	lie les fragments d'Okazaki entre eux hélicase ATP-dépendante protéine se liant à l'ADN monocaténaire	monomère 75 000 monomère 66 000 tétramère 74 000	environ 300 environ 50 environ 300

l'ATP). La primase stabilise la liaison entre le complexe DnaB-ATP et la molécule d'ADN monocaténaire.

Le site où se produit l'amorçage, et où il y a probablement formation du primosome, est connu chez plusieurs plasmides. En les comparant entre eux et en les mutant, nous devrions être capables de définir les caractéristiques importantes de la séquence d'ADN participant à l'amorçage.

Le primosome semble se déplacer, principalement, dans la direction opposée à celle de l'élongation, le long de la molécule parentale, c'est-à-dire de l'extrémité 5' vers l'extrémité 3'. La **Figure 14.10** montre que ceci pourrait être utilisable pour amorcer la synthèse des fragments d'Okazaki du brin secondaire d'un ADN bicaaténaire. En effet, au fur et à mesure que la fourche de réplication avance, il y a création d'une région monocaténaire en avant du primosome (sur son côté 3'). Après chaque amorçage, le primosome se déplace le long de cette région monocaténaire jusqu'au site d'amorçage du fragment d'Okazaki suivant. Par conséquent, le primo-

sosome se déplace dans le même sens que celui de la fourche de réplication, mais dans la direction opposée à la synthèse de l'ADN du brin secondaire.

Poursuivons l'analogie entre la réaction que nous avons caractérisée au niveau de l'ADN du phage $\phi X 174$ et la réplication de l'ADN double brin de *E. coli* en considérant comment, au niveau de ce dernier, est préparée la matrice simple brin en vue de la synthèse de la chaîne secondaire. Les protéines SSB se lient à l'ADN au fur et à mesure que la fourche de réplication avance, empêchant les deux brins parentaux de se réapparier afin que chacun d'entre eux soit apte à servir de matrice à la synthèse d'ADN. La protéine SSB est présente à un taux stœchiométrique à l'endroit de la fourche de réplication.

Comment les deux brins parentaux se désappariaient-ils initialement ? Les mutations dans le gène *rep*, qui code pour une ATPase, la protéine Rep, ralentissent le mouvement de la fourche de réplication. En présence d'ATP, l'action combinée des protéines Rep et SSB permet de séparer en ses constituants, un ADN double brin de

Tableau 14.5
D'autres protéines font aussi partie de l'appareil de réplication.

Protéine	Fonction	Produit (daltons)	Copies/Cellule
n' (= γ) n'' n	reconnaissance du site, ATPase (1/fourche) } préamorçage	monomère de 55 000 inconnu	environ 80 inconnu

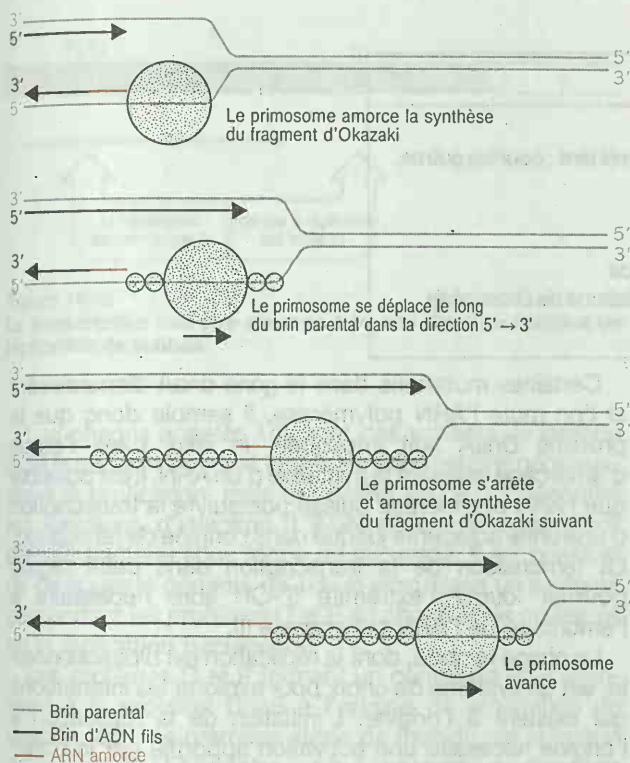


Figure 14.10

Le primosome suit la fourche de réplication en se déplaçant dans la direction 5'-3' le long du brin servant de matrice à la synthèse du brin secondaire, et amorce périodiquement la synthèse d'un fragment d'Okazaki.

ϕ X 174 dont un brin est coupé (cf. prochain paragraphe). Il est probable que la protéine Rep dissocie les chaînes qui sont ensuite stabilisées à l'état simple brin grâce aux protéines SSB.

Un modèle d'action de la protéine Rep est décrit dans la Figure 14.11. La protéine se lie à une région monocaténaire du double-brin d'ADN, et se déplace le long du brin où se trouve cette région dans une certaine direction. Quand la protéine atteint un segment bicaténaire, elle déplace l'extrémité 5'. Le désappariement d'une paire de bases conduit à l'hydrolyse de deux molécules d'ATP.

INITIATION DE LA RÉPLICATION AUX ORIGINES BICATÉNAIRES

Le commencement d'un cycle de réplication requiert deux types d'activité en plus des fonctions nécessaires

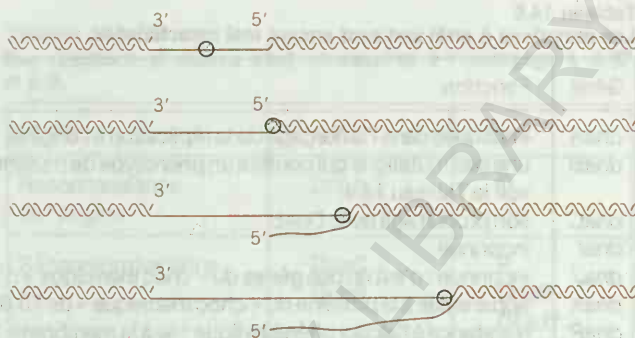


Figure 14.11

Il est possible que la protéine Rep avance le long d'un des brins d'ADN en déplaçant le brin complémentaire.

à la synthèse des fragments d'Okazaki. Premièrement, les deux brins de l'ADN doivent être séparés et deuxièmement, la synthèse du brin secondaire doit être amorcée. Le second événement est peut-être le résultat du même mécanisme que celui permettant l'amorçage de la synthèse des fragments d'Okazaki.

Les mutations à arrêt lent chez *E. coli* ont permis d'identifier plusieurs loci dont les produits pourraient participer aux événements se produisant à l'origine de réplication, mais très peu de choses sont connues quant à leurs activités. Sur tous les loci présentés dans le Tableau 14.6, seulement trois ont une fonction connue.

Les mutations dans les gènes *dnaB* et *dnaC* donnent lieu à des phénotypes aussi bien à arrêt lent qu'à arrêt rapide ; nous savons que les protéines codées par ces deux gènes jouent un rôle dans la réplication proprement dite (voir précédemment). Nous supposons qu'elles sont également requises au moment de l'initiation de la réplication.

La protéine DnaA se lie à une séquence consensus de 9 pb, qui est répétée quatre fois dans *oriC*. Sa liaison à l'ADN est hautement coopérative ; la protéine est en quantité telle qu'une région d'environ 100 pb au niveau d'une seule séquence consensus, et d'environ 250 pb sur *oriC* en est recouverte. En fait, il est possible que l'ADN s'enveloppe autour d'un noyau protéique de protéines DnaA. Toujours est-il qu'il est fort probable que la création d'une structure particulière soit une condition nécessaire à la séparation des deux brins d'ADN et à l'initiation de la réplication.

Les plasmides ayant intégré la séquence *oriC* permettent le développement de systèmes cellulaires de réplication à partir de cette origine. L'initiation de la réplication *in vitro*, à partir d'*oriC*, comporte deux étapes :

- Le préamorçage nécessite 6 protéines dont la liste figure dans le Tableau 14.7 ; elle permet la conversion

Tableau 14.6

Les mutations à arrêt lent sont encore mal caractérisées.

Gène	Fonction
<i>dnaA</i> <i>dnaB</i>	impliquée dans l'amorçage de la réplication à l'origine une des mutations qui confère un phénotype de mutant à arrêt lent ; pour les autres, voir le Tableau 14.4
<i>dnaC</i> <i>dnaI</i>	son produit agit avec DnaB inconnue
<i>dnaJ</i> <i>dnaK</i> <i>dnaP</i>	inconnue : c'est un des gènes du « choc thermique » d' <i>E. coli</i> apparentée à la protéine du « choc thermique » de 70 000 daltons de <i>Drosophila</i> résistance à l'alcool phénéthylrique ; lié à la membrane ?

d'une molécule circulaire superenroulée en une nouvelle forme partiellement désappariée. Cette réaction se produit en deux étapes ; il y a formation du complexe puis désappariement de l'ADN.

Ce sont les protéines DnaA, DnaB et DnaC qui sont les premières à former le complexe dont l'assemblage est stimulé par la protéine HU (une des protéines de type histone d'*E. coli*, voir Tableau 28.2). La protéine DnaB est présente à faible concentration à l'origine, contrairement à ce qui se passe aux fourches de réplication où là, de grandes quantités de protéines sont nécessaires. Sa fonction à l'origine est probablement catalytique ; elle semble en fait être l'hélicase qui permet la séparation des paires de bases. La gyrase, elle, sert de pivot à la dissociation des deux chaînes (cette réaction est détaillée dans le Chapitre 28), tandis que la protéine SSB se lie aux deux brins monocaténaux au fur et à mesure qu'ils se forment. Les deux étapes décrites, assemblage du complexe et désappariement des deux chaînes, consomment de l'énergie apportée sous forme d'ATP.

- L'amorçage nécessite la synthèse d'un ARN qui, grâce à son extrémité 3'-OH libre, va permettre à l'ARN polymérase III de commencer l'élongation. Selon la composition des systèmes *in vitro*, cette fonction est assurée soit par la DnaG primase, soit par l'ARN polymérase.

Tableau 14.7

Six protéines sont nécessaires pour le préamorçage à *oriC*.

Protéine	Fonctions
DnaA	protéine de 52 000 daltons qui se lie de façon coopérative à <i>oriC</i>
DnaB	1-2 hexamères forment un complexe à <i>oriC</i> : a l'activité hélicase
DnaC	forme un complexe avec DnaB à <i>oriC</i>
HU	stimule la formation du complexe
Gyrase	désapparie l'ADN afin d'engendrer des brins monocaténaux
SSB	stabilise l'ADN monocaténaire

Certaines mutations dans le gène *dnaA* disparaissent si l'on mute l'ARN polymérase, il semble donc que la protéine DnaA soit impliquée, *in vivo*, dans l'étape d'amorçage exigeant la synthèse d'un ARN. Il est possible que l'ARN polymérase puisse poursuivre la transcription d'une unité adjacente jusque dans l'origine de réplication. La terminaison de la transcription dans cette région pourrait fournir l'extrémité 3'-OH libre nécessaire à l'amorçage de l'ARN polymérase III.

Le phage lambda, dont la réplication est bidirectionnelle, est un système de choix pour explorer les interactions qui existent à l'origine. L'initiation de la réplication à l'origine nécessite une activation apportée par la transcription démarrant de *P_R*.

La Figure 14.12 montre une carte de la région concernée. Le besoin d'une synthèse d'ARN n'implique pas nécessairement que cet ARN fournisse l'amorce pour la synthèse du brin principal. Une alternative est que la synthèse d'ARN soit responsable de certains changements structuraux de la région.

L'initiation nécessite les produits des gènes *O* et *P*, ainsi que quelques protéines de l'hôte, comme, par exemple, les produits des gènes à arrêt lent. La protéine *O* du phage se lie à l'origine de réplication de lambda ; la protéine *P* du phage interagit avec la protéine *O* et d'autres protéines bactériennes. L'origine est située dans le gène *O* ; de ce fait, la protéine *O* agit près de son site de synthèse.

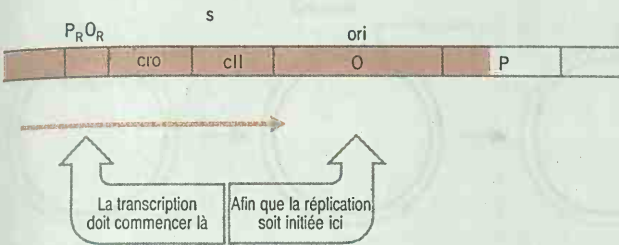


Figure 14.12 La transcription doit être amorcée à P_R afin d'activer l'origine de réplication de lambda.

Les phages appelés λdv sont des variants de lambda dont le génome contient toutes les informations nécessaires à la réplication, mais est délété en ce qui concerne les fonctions d'infection. L'ADN de ces phages survit dans la bactérie sous forme de plasmide. Il est possible de répliquer le génome de λdv *in vitro*, dans un système comportant les protéines O et P du phage et les protéines de la réplication bactérienne.

Les protéines O et P forment un complexe en s'associant à la protéine DnaB à l'origine de réplication de lambda, *ori λ* . La première étape de fixation est la liaison de la protéine O à l'origine, ce qui engendre une structure grossièrement sphérique d'un diamètre d'environ 11 nm appelée parfois O-some. L'O-some contient environ 100 pb, ce qui représente environ 60 000 daltons d'ADN. Il y a quatre sites de fixation pour la protéine O dont le poids moléculaire est de 34 000 daltons. Chacun de ces sites est palindromique, ce qui permet probablement la liaison d'un dimère symétrique de protéines O. L'ADN subit un changement structural durant la formation du complexe ; il est possible qu'il se courbe sur l'O-some.

Une fois les protéines P phagique et DnaB bactérienne ajoutées, le complexe s'agrandit et devient asymétrique. Il contient, et plus de protéines, et plus d'ADN (environ 160pb au total). Ensuite, les autres protéines impliquées dans les fonctions à arrêt lent (DnaJ et DnaK) vont se joindre au complexe et former le complexe de préinitiation. Enfin, la primase initie la réplication.

Les similarités entre les réactions d'initiation aux loci *oriC* et *ori λ* sont récapitulées dans le **Tableau 14.8**. Dans les deux cas, les mêmes étapes sont requises pour l'initiation, qui nécessite des protéines diverses. Il est à noter que l'attribution des différentes fonctions à telle ou telle protéine est présomptive. La première étape est sans doute la reconnaissance de l'origine par une protéine qui se lie à l'ADN (la protéine DnaA pour *oriC* et la protéine O pour *ori λ*). Un complexe, dont la protéine DnaB est un composant commun probablement impliqué dans la

Tableau 14.8 Des réactions similaires sont nécessaires à l'amorçage à *oriC* et *ori λ* .

Étape	<i>oriC</i>	<i>oriλ</i>
Reconnaissance de l'origine	DnaA	O
Séparation des brins	DnaC DnaB	P DnaB DnaJ? DnaK?
Désappariement	Gyrase SSB	? ?
Amorçage	ARN pol	ARN pol

séparation des deux chaînes, sert ensuite à la réaction de préamorçage. Enfin, la synthèse d'un ARN amorce va permettre à la réplication de commencer.

La synthèse du brin principal à partir de l'*oriC* ou à partir de l'*ori λ* dépend de la présence de l'activité de synthèse d'amorce (assurée par l'ARN polymérase, parfois en association avec la protéine DnaG), tout comme l'initiation des fragments d'Okazaki du brin secondaire. La seule différence est que l'élongation du brin principal se fait de façon continue, alors que la synthèse du brin secondaire est discontinue. Néanmoins, certains systèmes ont recours à un autre moyen pour initier la réplication à l'origine. Dans ces cas-là, un des brins parentaux est coupé, ce qui engendre une extrémité 3'-OH libre qui est alors utilisée en tant qu'amorce comme le montre la **Figure 14.13**. Par la suite, la liaison covalente liant l'un des brins parentaux à l'un des brins fils est rompue.

Ce mécanisme — coupure puis élongation d'un des brins parentaux — est utilisé par certains phages dont la réplication s'effectue sur le mode du cercle roulant, précédemment illustré sur la Figure 13.10. Ceci a été mis en évidence dans le cas de l'ADN RF de $\phi X174$. Il y a moins de fonctions enzymatiques impliquées dans la réplication par le cercle roulant que lors de la production d'ADN RF à partir du brin viral originel. Les formes circulaires superenroulées de l'ADN RF de $\phi X174$ peuvent être répliquées *in vitro* grâce à un système de réplication comportant le produit du gène A du phage, la protéine Rep de l'hôte, SSB et l'ADN polymérase III.

La protéine A coupe le brin (+) viral à un site spécifique de la molécule d'ADN bicaténaire, qui est défini comme étant l'origine de réplication de RF. L'ADN ne peut être copié que s'il est sous forme superenroulée. La protéine

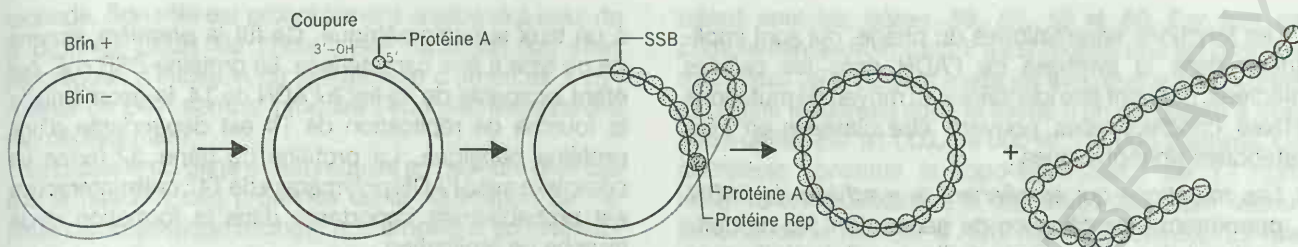


Figure 14.14

L'ADN RF de ϕ X174 peut être converti en deux molécules monocaténares grâce à l'effet combiné de trois fonctions : coupure par la protéine A, désappariement des deux brins par la protéine Rep et stabilisation des structures monocaténares par la protéine SSB.

Après la coupure de l'ADN, le brin (+) désapparié est libéré sous la forme d'une molécule linéaire qui peut être recircularisée. La protéine A est impliquée dans cette recircularisation. En effet, la liaison des extrémités 5' et 3' du brin (+) est concomitante de la réaction permettant à la protéine A de couper le brin (+) et de rester attachée à l'extrémité libre engendrée par cette coupure.

La protéine A a la propriété peu commune d'agir en *cis in vivo*, ce qui signifie qu'elle ne peut se lier qu'à l'ADN du génome qui en a permis sa synthèse. (Cette propriété n'est pas conservée *in vitro*, la protéine peut se lier à n'importe quel ADN dans un système acellulaire.) Nous ne connaissons pas le mécanisme de cette reconnaissance. Néanmoins, l'étude *in vitro* de la protéine a permis de comprendre comment elle restait toujours associée au brin (-) parental.

Lors de la morphogénèse du phage, le cycle présenté sur la Figure 14.15 est légèrement modifié. Le brin (+) désapparié n'est pas recouvert de protéines SSB (puis

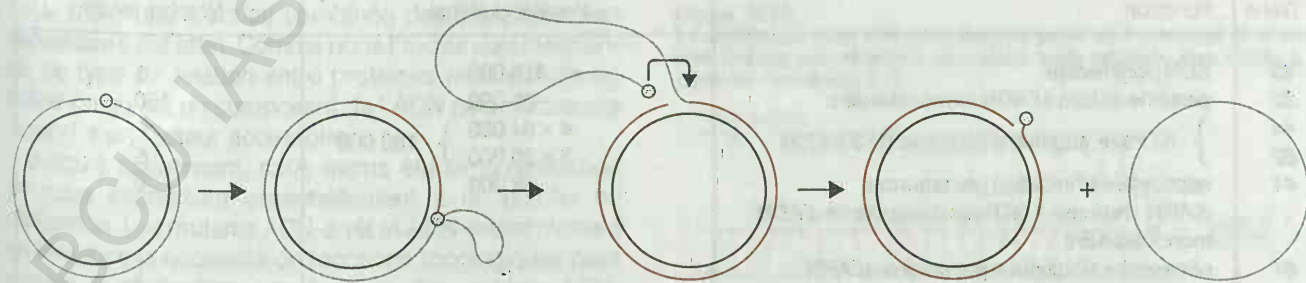
réengagé dans le cycle de réplication), mais il entre dans le virion phagique.

L'APPAREIL DE RÉPLICATION DU PHAGE T4

Quand le phage T4 infecte une cellule, il pourvoit lui-même à certaines fonctions enzymatiques qui remplacent ou augmentent celles de l'hôte. L'ADN de l'hôte se dégradant, le phage a peu confiance dans l'expression des fonctions enzymatiques de l'hôte sauf en ce qui concerne l'utilisation des enzymes présentes en grande quantité dans la bactérie. La dégradation de l'ADN bactérien est importante car elle permet la libération de nucléotides qui seront utilisés dans la synthèse de l'ADN du phage. (La composition en base de l'ADN du phage diffère de celle de l'ADN cellulaire en ce qu'il est constitué d'hydroxyméthylcytosine à la place de la cytosine habituelle.)

Figure 14.15

L'ADN RF de ϕ X174 peut être employé comme modèle pour la synthèse de molécules virales circulaires et monocaténares. La protéine A reste attachée au même génome au cours d'un nombre illimité de révolutions, en coupant à l'origine du brin viral (+) après chaque tour et en se déplaçant sur la nouvelle extrémité 5'. Au même moment, le brin viral relargué est circularisé.



Les fonctions enzymatiques du phage, qui sont impliquées dans la synthèse de l'ADN dans les cellules infectées, peuvent être identifiées au moyen de mutations létales conditionnelles pouvant être classées en trois catégories phénotypiques :

- Les mutations qui *empêchent la synthèse de l'ADN*, permettent l'identification de gènes dont les produits sont soit des composants de l'appareil de réplication, soit des protéines impliquées dans la synthèse des nucléotides (plus particulièrement l'hydroxyméthylcytosine).
- Les mutations qui *retardent la synthèse de l'ADN* concernent des gènes impliqués dans l'initiation de la réplication.
- Les mutations qui *arrêtent la synthèse d'ADN* après qu'elle ait été initiée, portent sur les gènes codant pour des fonctions de régulation, l'ADN ligase ou des enzymes impliquées dans la dégradation de l'ADN de l'hôte.
- Des gènes « non essentiels » sont également impliqués dans la réplication, comme, par exemple, les gènes qui interviennent dans la glycosylation de l'hydroxyméthylcytosine.

La synthèse de l'ADN de T4 est catalysée par un complexe multi-enzymatique composé des produits de sept gènes essentiels. Chacun de ces produits a été purifié par test de complémentation *in vitro*. Le complexe est actif *in vitro* sur des ADN d'origines différentes. Il agit à une vitesse comparable à sa vitesse *in vivo*, et sa fréquence d'erreur est d'environ 10^{-5} - 10^{-6} (la fréquence d'erreur *in vivo* étant d'environ 2×10^{-8}). *In vitro*, le CTP est utilisé aussi efficacement que l'hydroxyméthyl-CTP.

Les protéines du complexe de réplication de T4 sont décrites dans le **Tableau 14.9**.

La protéine du gène 32 se lie de façon hautement coopérative à l'ADN monocaténaire, où elle est requise

à un taux stœchiométrique. Ce fut la première enzyme de ce type à être caractérisée. La protéine SSB d'*E. coli* étant incapable de se lier à l'ADN de T4, la géométrie de la fourche de réplication de T4 est dépendante d'une protéine phagique. La protéine du gène 32 forme un complexe avec l'ADN polymérase de T4 ; cette interaction est probablement importante dans la formation de la fourche de réplication.

Le système de réplication de T4 est dépendant de l'activité de synthèse d'ARN amorce. La nature de l'amorce est toutefois différente de celle utilisée par *E. coli*. Quand c'est l'ADN monocaténaire de T4 qui est à répliquer, les produits des gènes 41 et 61 agissent ensemble afin de synthétiser de courtes amorces. Ce sont des pentaribonucléotides dont la séquence consensus est pppApCpNpNpNp. Leur longueur est toujours limitée à cinq bases ; les deux premiers nucléotides de la séquence sont spécifiques alors que les trois derniers sont représentés indifféremment par n'importe quelle base.

Les amorces sont ensuite allongées en chaînes d'ADN si toutefois l'appareil de réplication est présent dans son entier. Les amorces sont synthétisées à partir de sites spécifiques. Ces sites apparaissent sur l'ADN plus fréquemment que ceux utilisés pour l'initiation des fragments d'Okazaki (qui apparaissent toutes les 2 000 bases environ). En réalité, il est probable que la fourche de réplication passe sans s'arrêter sur une série de sites potentiels de synthèse d'amorce, et que ce soit la distance parcourue depuis la dernière initiation qui détermine la position de la prochaine initiation.

La protéine du gène 41 est une hélicase (capable de désappairer la double hélice). Elle a également une activité de GTPase stimulée par l'ADN monocaténaire. Quand la protéine se lie à un brin d'ADN, elle est capable de se déplacer à partir de son site de liaison initial et d'avancer à une vitesse d'environ 400 nucléotides par

Tableau 14.9

Les composants de l'appareil de réplication de T4 sont définis.

Gène	Fonction	Produit (daltons)	Nombre relatif
43	ADN polymérase	110 000	1
32	protéine se liant à l'ADN monocaténaire	35 000	150
44	} ATPase, augmente la vitesse de 3 à 4 fois	$4 \times 34\ 000$	180 000
62		$2 \times 20\ 000$	
41		58 000	
	nécessaire à l'initiation par amorces d'ARN : hélicase 1 GTPase dépendante d'ADN monocaténaire		5
61	nécessaire à l'initiation par amorce d'ARN		20
45	aide l'ADN polymérase	$2 \times 27\ 000$	1
			10

seconde. Son rôle est probablement analogue à celui de la protéine DnaB de l'hôte ; elle reconnaît les sites périodiques d'initiation de la synthèse d'amorces. L'hydrolyse de GTP semble produire l'énergie nécessaire à son déplacement.

La protéine du gène 61 est requise en moindre quantité que la majorité des protéines de la réplication de T4. Cette caractéristique a longtemps empêché son identification (en effet, sa présence en faible quantité, en tant que contaminant de la préparation de la protéine du gène 32, passait inaperçue dans les systèmes de réplication !). Il est possible que ce soit une primase, analogue à la protéine DnaG d'*E. coli*. Elle est présente à environ 10 copies par cellule.

L'ADN polymérase du gène 43 possède l'activité de synthèse 5'-3' associée à l'activité exonucléasique de « correction sur épreuve » 3'-5'. Les trois protéines restantes sont appelées les « protéines secondaires de la polymérase ». Le produit du gène 45 est un dimère. Les produits des gènes 44 et 62 forment un complexe à activité ATPasique qui augmente la vitesse du mouvement de l'ADN polymérase de 3-4 fois (de 250 nucléotides/seconde à environ 800 nucléotides/seconde, c'est-à-dire très proche de la vitesse de la réplication *in vivo* qui est d'environ 1 000 nucléotides/seconde), et ce, que l'ADN à copier soit monocaténaire ou bicaténaire.

Si l'ADN polymérase de T4 doit copier un ADN monocaténaire, sa vitesse de progression est inégale. En effet, l'enzyme se déplace rapidement le long des régions monocaténaires, mais ralentit de façon notable quand elle parcourt des régions appariées au sein du brin. Les protéines accessoires aident l'ADN polymérase à passer ces « barrages ». Leur fonction d'augmentation de vitesse de réplication n'est peut-être en fait qu'un moyen de ne pas laisser ralentir l'ADN polymérase lorsque celle-ci rencontre des passages difficiles.

La présence de ces protéines augmente également l'affinité de l'ADN polymérase pour l'ADN, ainsi que sa capacité à rester sur la même molécule sans s'en dissocier. Ces protéines doivent, en quelque sorte, « agrafer » l'ADN polymérase à l'ADN en le maintenant étroitement sur la molécule. L'action combinée des 3 protéines est nécessaire à cet effet. Comme nous l'avons déjà mentionné, ce type de relation entre protéines rend difficile de définir ce qu'est un composant de l'ADN polymérase par rapport à un facteur accessoire.

Jusqu'à maintenant, nous avons étudié la réplication en nous intéressant essentiellement à la fourche de réplication. Les mutants ADN-arrêt et ADN-retard mettent en évidence la nécessité de fonctions topologiques pour assurer la réplication. Les 4 gènes des mutants ADN-

retard sont les gènes 39, 52, 58 et 60. Par test de complémentation *in vitro*, il a été montré que les produits des gènes 39 et 52 sont associés avec une autre protéine (d'origine inconnue) sous la forme d'un complexe à trois sous-unités (de 51 000, 64 000 et 110 000 daltons). Ce complexe constitue la topo-isomérase II de T4 dont l'activité est nécessaire à la disparition des super-enroulements de la matrice d'ADN (voir Chapitre 28). Le rôle de cette protéine est essentiel pour la réplication, ce qui suggère que l'ADN de T4 n'est pas maintenu sous forme linéaire ne possédant pas d'extrémité libre. La topo-isomérase est peut-être également utilisée dans le désappariement de l'ADN en avant de la fourche de réplication.

LE PROBLÈME DES RÉPLICONS LINÉAIRES

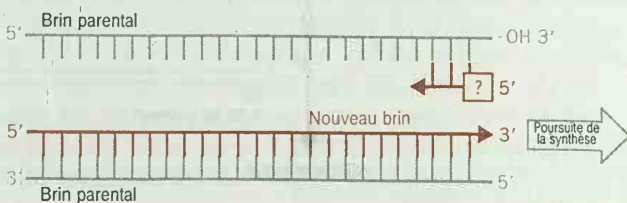
Toutes les ADN ou ARN polymérases connues n'agissent qu'en se déplaçant dans la direction 5'-3'. Cette propriété pose un problème essentiel en ce qui concerne la synthèse d'ADN aux extrémités 5' d'un réplicon linéaire. Considérons les deux brins parentaux décrits sur la Figure 14.16. Sur le brin inférieur, nous voyons que la réplication ne pose aucun problème à partir du moment où elle a démarré : le brin inférieur agit comme matrice pour la synthèse du brin fils, synthèse qui va se poursuivre jusqu'à l'extrémité terminale 3', où probablement la polymérase se décroche. Sur le brin supérieur, c'est la situation au début de la réplication qui est mise en relief. En effet, afin de synthétiser le brin complémentaire, un événement d'initiation doit se produire juste au niveau de la toute première base en 5'.

Nous ne savons pas si une initiation à l'extrémité terminale est possible. Nous pensons plutôt que l'ADN polymérase se lie à un site qui lui permet d'englober le lieu où la base va être incorporée. Quatre solutions peuvent être proposées qui permettent la copie de l'extrémité terminale :

- Le problème est circonvenu si le réplicon passe de la forme linéaire à la forme circulaire ou s'il se concatémérise, comme c'est le cas pour les phages T4 et lambda.

Figure 14.16

La réplication peut être poursuivie à partir de l'extrémité 3' d'un brin linéaire nouvellement synthétisé mais peut-elle être initiée à partir de l'extrémité 5' ?

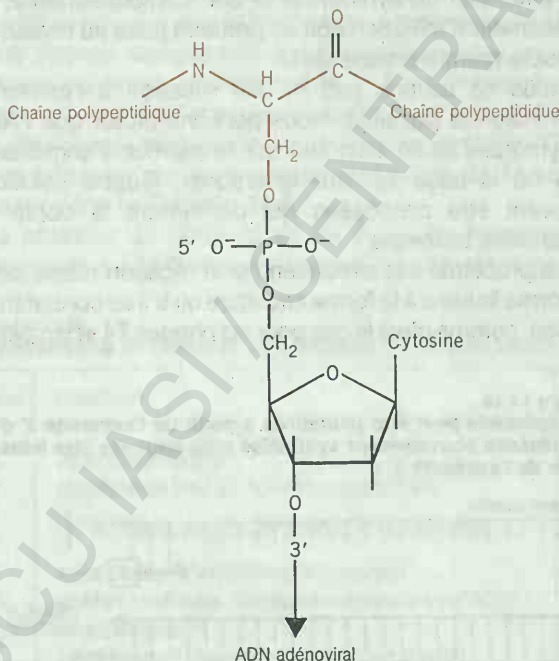


- L'ADN peut adopter une structure inhabituelle — par exemple, une boucle en épingle à cheveux — à l'extrémité terminale, évitant ainsi la présence d'extrémités libres. C'est le cas de la réplication de l'ADN mitochondrial linéaire de *Paramecium*.
- Au lieu d'être déterminée de façon précise, la position de l'extrémité est variable. Les télomères des chromosomes des eucaryotes inférieurs adoptent parfois cette solution en modifiant le nombre de copies d'une courte séquence répétée se trouvant à leur extrémité (voir Chapitre 25). S'il existe un mécanisme qui permet d'ajouter ou d'ôter cette séquence répétée, il devient alors inutile de répliquer l'ADN jusqu'à son extrémité terminale.
- La solution la plus directe est de posséder une protéine qui intervient afin de rendre cette réaction d'initiation possible. Certains acides nucléiques linéaires viraux codent pour des protéines qui se *lient de façon covalente à leurs extrémités 5' terminales*. Les exemples les mieux caractérisés sont ceux de l'ADN de l'adénovirus, du phage $\phi 29$ et de l'ARN du poliovirus.

L'ADN de l'adénovirus est une molécule linéaire bicaaténaire dont les deux extrémités 5' sont liées de façon

Figure 14.17

Les extrémités phosphate 5' terminales de l'ADN adénoviral sont liées de façon covalente à une sérine de la protéine de 55 000 daltons.



covalente à une protéine de 55 000 daltons. Comme l'indique la **Figure 14.17**, une liaison phosphodiester unit l'ADN à la sérine. Les extrémités 5' du phage $\phi 29$ sont liées de la même façon à une protéine de 27 000 daltons. Dans le cas de l'ARN monocaténaire du poliovirus, la protéine VPg de 22 acides aminés est liée à l'extrémité 5' terminale par le groupe hydroxyle de la tyrosine. Dans tous les cas, la protéine d'attache est codée par le virus et est impliquée dans la réplication.

L'ADN de l'adénovirus et de $\phi 29$ se réplique, à partir des extrémités 5', par le mécanisme de **déplacement de brin** qui est illustré sur la **Figure 14.18**. Des événements identiques, mais non simultanés, se produisent aux deux extrémités de la molécule. La synthèse d'une nouvelle chaîne est initiée et déplace le brin homologue qui était précédemment apparié. Selon les cas, il y a une seule fourche de réplication qui se déplace jusqu'à l'autre extrémité de la molécule, ou alors il existe deux fourches de réplication qui avancent à partir de chacune des extrémités et se rencontrent quelque part au milieu de la molécule. Quel que soit le modèle, il en résulte la séparation des deux chaînes.

La protéine présente sur les chaînes d'ADN en cours de réplication de l'adénovirus est une protéine de 80 000 daltons, donc de poids moléculaire supérieur à la protéine rencontrée sur l'ADN isolé à partir de virions. Ces deux protéines sont apparentées, ce qui suggère que la protéine de 80 000 daltons, impliquée dans la réplication de l'ADN, est clivée au cours de la maturation du virus, en une protéine de plus faible poids moléculaire.

Comment la liaison d'une protéine à l'extrémité 5' de l'acide nucléique permet-elle l'initiation ? La protéine de 80 000 daltons de l'adénovirus est capable de se lier de façon covalente au dCTP dans des extraits de cellules infectées par le virus. Elle est associée à une autre protéine, une ADN polymérase de 140 000 daltons. Ces observations ont suggéré le modèle décrit sur la **Figure 14.19**.

La protéine de 80 000 daltons porte une cytidine avec un groupe 3'-OH libre. Elle forme, avec l'ADN polymérase, un complexe qui aborde l'extrémité de l'ADN adénoviral. Le complexe déplace l'extrémité 5' qu'elle peut reconnaître grâce à la liaison à ce niveau de la version de 55 000 daltons de la protéine. L'extrémité 3'-OH est ensuite présentée à un nucléotide qui s'apparie au brin parental sous la direction de l'ADN polymérase. La réplication de $\phi 29$ suit probablement le même mécanisme.

Cette série de réactions doit être coordonnée car la protéine de 80 000 daltons ne peut se lier au CTP qu'en présence de l'ADN polymérase et de l'ADN adénoviral.

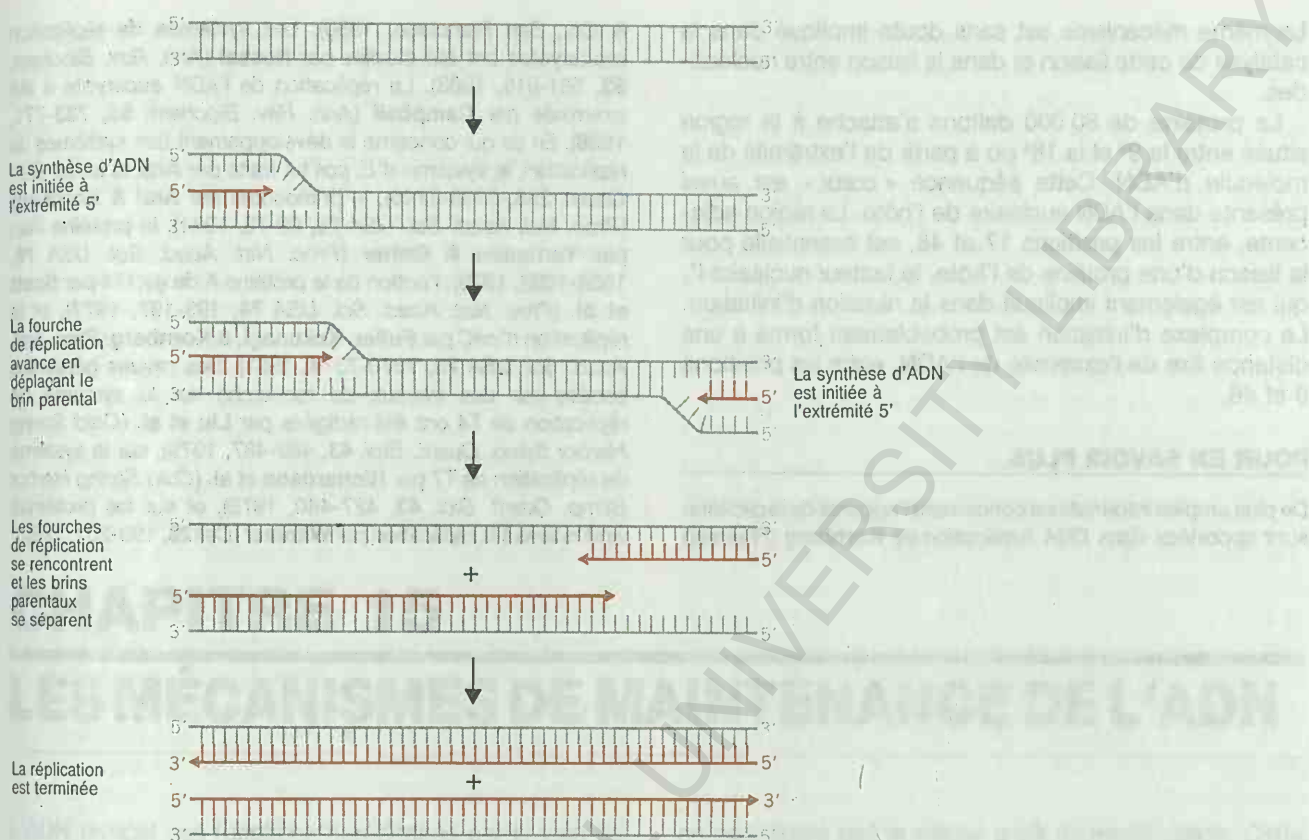
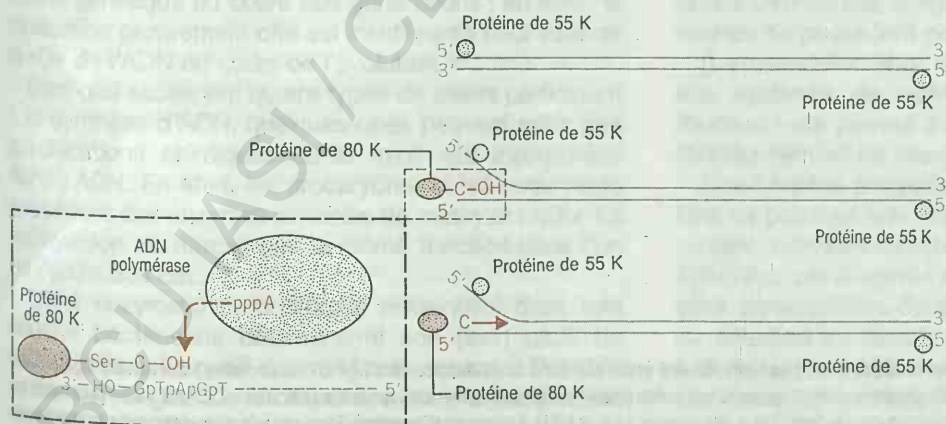


Figure 14.18
La réplication de l'ADN adénoviral est initiée indépendamment aux deux extrémités de la molécule et avance par déplacement de brin.

Figure 14.19
La protéine adénovirale de 80 000 daltons déplace l'extrémité 5' de l'ADN (qui est liée à la protéine de 55 000 daltons) et apporte un CTP afin de permettre l'initiation de la synthèse d'un nouveau brin d'ADN.



Le même mécanisme est sans doute impliqué dans la catalyse de cette liaison et dans la liaison entre nucléotides.

La protéine de 80 000 daltons s'attache à la région située entre la 9^e et la 18^e pb à partir de l'extrémité de la molécule d'ADN. Cette séquence « cœur » est aussi présente dans l'ADN nucléaire de l'hôte. La région adjacente, entre les positions 17 et 48, est essentielle pour la liaison d'une protéine de l'hôte, le facteur nucléaire I*, qui est également impliqué dans la réaction d'initiation. Le complexe d'initiation est probablement formé à une distance fixe de l'extrémité de l'ADN, entre les positions 9 et 48.

POUR EN SAVOIR PLUS

De plus amples informations concernant l'appareil de réplication sont apportées dans *DNA Replication* de Kornberg (Freeman

& Co., San Francisco, 1980). Les systèmes de réplication procaryotes ont été étudiés par Nossal (*Ann. Rev. Biochem.* **53**, 581-615, 1983). La réplication de l'ADN eucaryote a été examinée par Campbell (*Ann. Rev. Biochem.* **55**, 733-771, 1986). En ce qui concerne le développement des systèmes de réplication, le système d'*E. coli* fut traité par Arai et al. (*J. Biol. Chem.* **256**, 5239-5246), le primosome par Arai & Kornberg (*Proc. Nat. Acad. Sci. USA* **78**, 69-73, 1981), la protéine Rep par Yarranton & Gefter (*Proc. Nat. Acad. Sci. USA* **76**, 1658-1662, 1979), l'action de la protéine A de ϕ X174 par Scott et al. (*Proc. Nat. Acad. Sci. USA* **74**, 193-197, 1977), et la réplication d'*oriC* par Fuller, Kakunagi, & Kornberg (*Proc. Nat. Acad. Sci. USA* **78**, 7370-7374, 1981). Des revues générales basées sur des travaux de recherche sur le système de réplication de T4 ont été rédigées par Liu et al. (*Cold Spring Harbor Symp. Quant. Biol.* **43**, 469-487, 1979), sur le système de réplication de T7 par Richardson et al. (*Cold Spring Harbor Symp. Quant. Biol.* **43**, 427-440, 1978), et sur les protéines virales liées à la réplication par Wimmer (*Cell* **28**, 199-201, 1982).

* N. d. T. : Le facteur nucléaire I (NFI en abrégé) n'est pas uniquement impliqué dans la réplication. Il s'agit d'une protéine intervenant également dans la transcription d'un grand nombre de gènes eucaryotiques. C'est cette protéine NFI qui se lie notamment aux boîtes « CCAAT » des promoteurs des gènes alpha et bêta globine. Un troisième type de fonction a été suggéré pour NFI : il servirait d'intermédiaire pour la stimulation de promoteurs par le facteur de croissance TGF bêta.

CHAPITRE 15

LES MÉCANISMES DE MAINTENANCE DE L'ADN

L'ADN remplit ses fonctions héréditaires par la voie de la réplication et de la transcription. A chacun de ces mécanismes sont associées de nombreuses activités enzymatiques et régulatrices ainsi que les polymérases appropriées. En plus de ces systèmes, il en existe d'autres impliquant des enzymes qui interagissent avec l'ADN afin de modifier sa structure et afin de réparer les dommages qu'il a éventuellement subi. Ces mécanismes sont essentiels pour permettre la conservation du patrimoine génétique au cours des générations ; en effet, la réplication proprement dite est insuffisante pour assurer le rôle de l'ADN au cours de l'évolution.

Bien que seulement quatre types de bases participent à la synthèse d'ADN, quelques-unes peuvent subir des modifications chimiques après avoir été incorporées dans l'ADN. En effet, les procaryotes et les eucaryotes possèdent des enzymes capables de méthyler l'ADN. La méthylation ne remplit pas la même fonction dans l'un ou l'autre des cas.

L(es) enzyme(s) particulière(s) présente(s) dans une souche bactérienne détermine(nt) son (leur) profil de **modification**. Le profil de modification permet l'identification de l'origine de l'ADN. Grâce à lui, la bactérie est capable de distinguer son propre ADN d'un ADN étranger

ne possédant pas le même profil de modification. Cette différence permet l'attaque de l'ADN étranger par des **enzymes de restriction** qui reconnaissent, à des sites précis, l'absence de groupements méthyle.

Ces systèmes sont appelés systèmes de **restriction et modification**. Ils sont largement répandus dans les bactéries. Ils agissent dans un but xénophobique c'est-à-dire qu'ils protègent l'ADN de l'hôte contre toute contamination par des séquences d'origine étrangère (leur présence n'est pas obligatoire ; certaines souches bactériennes ne possèdent pas de systèmes de restriction).

(La méthylation chez les eucaryotes n'est pas associée aux systèmes de restriction ; elle a une toute autre fonction ; elle permet à la cellule de distinguer l'état de fonctionnement de ses gènes, voir Chapitre 27.)

Une bactérie possède plusieurs systèmes lui permettant de protéger son ADN contre les conséquences des erreurs commises lors de la réplication ou des dommages subis à cause d'agents extérieurs. Tout événement qui a pour conséquence de dévier irréversiblement l'ADN de sa structure en double hélice, est inadmissible pour la cellule. Ce peut être une mutation ponctuelle remplaçant une base par une autre et créant ainsi un appariement de bases incompatible avec les règles énoncées par

Watson et Crick, ou encore un changement structural qui encombre l'ADN ou lie deux bases ensemble.

Les sites où se produisent des mutations sont reconnus spécifiquement par des nucléases qui excisent la région endommagée de l'ADN ; d'autres enzymes se chargent ensuite de synthétiser la séquence manquante. Ensemble, ces différentes activités forment le **système de réparation**. En plus de ce système de réparation par excision/resynthèse, il en existe d'autres qui limitent les conséquences de la réplication d'un ADN endommagé. Ces **systèmes de rétablissement** sont liés aux systèmes impliqués dans la recombinaison génétique.

Ces systèmes variés mettent tous en œuvre des enzymes remarquables dans leur capacité à reconnaître des séquences ou des structures particulières de l'ADN. Les enzymes de restriction se lient à l'ADN au niveau de séquences cibles spécifiques. Cette propriété apporte une nouvelle dimension à la nature des interactions acides nucléiques/protéines. Certaines de ces enzymes se lient à l'ADN en un site, mais clivent à un autre site très éloigné, ce qui met en évidence la capacité des protéines à se déplacer le long du double-brin d'ADN. D'autres enzymes de réparation reconnaissent des sites dans l'ADN à cause de leur distorsion par rapport à la structure normale de la double hélice. Les enzymes impliquées dans la recombinaison sont capables de lier deux molécules d'ADN afin de permettre leur appariement.

Toutes ces activités nous apprennent beaucoup quant aux interactions protéines/ADN qu'elles soient dépendantes de la séquence ou de la structure de l'ADN, dont l'existence est sans nul doute un point essentiel de la biologie du gène. Grâce à elles, nous commençons à avoir une connaissance générale des systèmes nécessaires à la préservation et à la protection de l'ADN contre les agressions de l'environnement ou contre les erreurs d'autres systèmes cellulaires.

MODIFICATION ET RESTRICTION

Les endonucléases de restriction bactériennes reconnaissent toutes de courtes séquences d'ADN comme sites de liaison. La liaison est suivie du clivage de l'ADN à l'endroit précis du site de reconnaissance ou autre part, selon l'enzyme. Dans les Chapitres 4 et 16, nous avons déjà discuté de l'intérêt des enzymes de restriction lors de la cartographie ou de la reconstruction d'ADN *in vitro*. Nous allons maintenant nous intéresser à la fonction de ces enzymes *in vivo*.

En plus de son (ou ses) activité(s) de restriction, une

souche bactérienne possède une activité d'ADN méthylase dont la *spécificité de séquence est identique* à celle de l'activité de restriction. La méthylase ajoute des groupements méthyls (aux résidus adénine et cytosine) à la séquence cible reconnue par l'enzyme de restriction. La méthylation rend la séquence cible résistante à la restriction. La méthylase joue donc un rôle de protection contre le clivage par l'enzyme de restriction.

Les séquences cibles de l'ADN bactérien résidant sont méthylées, ce qui rend l'ADN indifférent aux attaques des enzymes de restriction, comme l'indique la **Figure 15.1**. Si cela n'était pas le cas, l'enzyme de restriction, en le dégradant, détruirait l'ADN de la cellule dans laquelle elle réside. Cette protection n'existe pas pour un ADN étranger qui réussit à pénétrer dans la cellule. En effet, les ADN étrangers à la cellule ne sont pas modifiés quant aux séquences cibles des enzymes de restriction, ce qui les rend sensibles à leurs attaques. La combinaison restriction/modification permet à la cellule de distinguer l'ADN étranger du sien, qu'elle a protégé.

Dans ces conditions, est considéré comme ADN étranger tout ADN issu d'une autre souche bactérienne *ne possédant pas la même activité de modification et de restriction*. La différence d'activité n'est pas liée aux différents types d'ADN bactérien, plasmidique ou phagique. En effet, l'ADN résidant d'une certaine souche bactérienne a le même profil de modification qu'un ADN laissé intact après passage dans cette même souche. Le système de restriction et de modification n'est pas un obstacle à l'échange d'ADN entre bactéries issues d'une même souche.

De fait, ces systèmes furent initialement découverts grâce à l'infection de souches bactériennes par de l'ADN phagique. En effet, ce dernier, après avoir été relargué de la souche bactérienne, peut aller infecter une autre bactérie de la même souche car il a le même profil de modification que l'ADN de l'hôte. Il ne peut cependant pas se développer dans une autre souche bactérienne car après l'avoir infectée, il est immédiatement attaqué par son activité de restriction. Le phage est « restreint » à une souche bactérienne.

La restriction n'est pas absolue. Quelques phages infectieux lui échappent soit par mutation dans les séquences cibles, soit par négligence du système de l'hôte. Dans le dernier cas, ils acquièrent le profil de modification de leur nouvel hôte.

Des plasmides peuvent être impliqués dans la spécificité de la souche bactérienne. En effet, certains plasmides (et phages) possèdent des gènes codant pour les systèmes de restriction et de modification ; par conséquent,

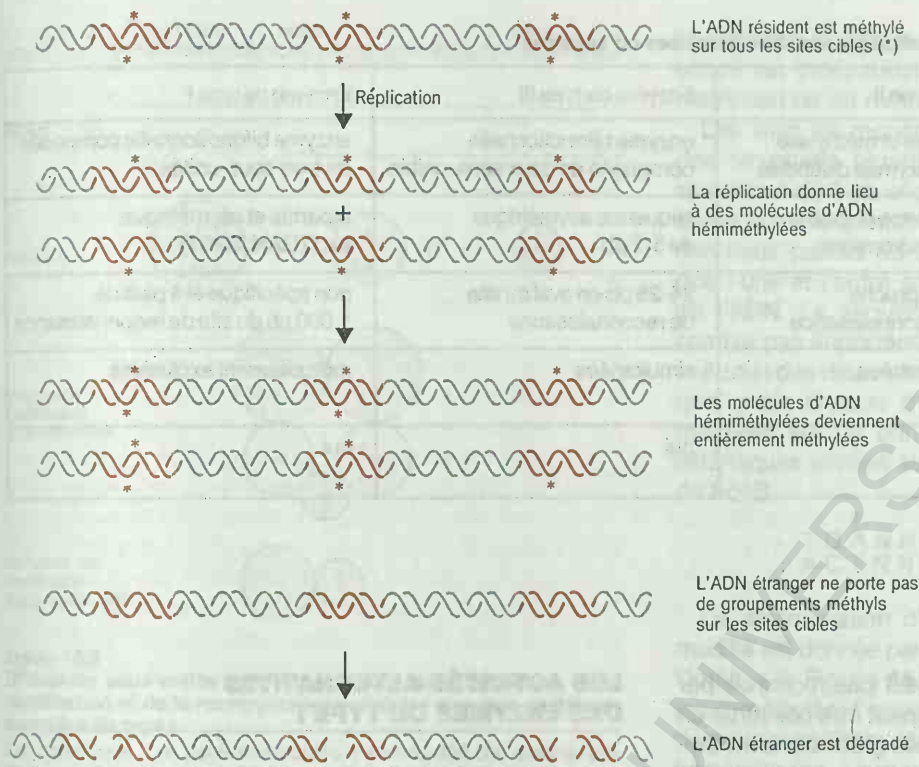


Figure 15.1
Les sites méthylés se perpétuent indéfiniment au cours des générations, ils sont protégés de la restriction ; les sites non méthylés sont coupés.

une bactérie portant un de ces éléments sera de souche différente d'une bactérie ne portant pas ce même élément. L'existence de tels systèmes sert de défense contre la mobilité des séquences bactériennes, résultant du passage de plasmides et de phages dans les bactéries.

Les enzymes de restriction sont classées en deux grandes catégories qui peuvent ensuite être sous-divisées. Leurs propriétés sont données dans le **Tableau 15.1**.

La première catégorie regroupe les enzymes ayant des activités de **type II**, ce qui comprend les enzymes décrites précédemment. Un système de type II existe dans une souche bactérienne sur trois. Les enzymes de type II ne portent que l'activité de restriction. Une autre enzyme est responsable de la méthylation de la séquence cible.

Les structures protéiques sont relativement simples.

L'exemple le mieux caractérisé est celui du système EcoRI, où l'enzyme de restriction est un dimère de sous-unités identiques et où la méthylase est un monomère (la reconnaissance, par ces deux enzymes, d'une séquence cible commune, peut être le résultat soit de la duplication d'une partie de la protéine qui reconnaît l'ADN, soit d'une évolution convergente).

Les séquences cibles sont souvent des palindromes de 4-6 pb. Étant donné la symétrie, les bases pouvant être méthylées sont présentes sur les deux brins d'ADN. Ainsi, une séquence cible *peut être totalement méthylée* (les deux brins sont modifiés), *hémiméthylée* (seul l'un des brins est méthylé) ou non méthylée :

- Une séquence totalement méthylée n'est la cible d'aucune restriction ou modification.

Tableau 15.1

Les activités de restriction et de méthylation peuvent être associées ou séparées.

	Enzyme de type II	Enzyme de type III	Enzyme de type I
Structure protéique	endonucléase et méthylase sont deux enzymes distinctes	enzyme bifonctionnelle composée de deux sous-unités	enzyme bifonctionnelle composée de trois sous-unités
Site de reconnaissance	courte séquence (4-6 pb) souvent palindromique	séquence asymétrique de 5-7 pb	bipartite et asymétrique ex. (TGAN ₈ TGCT)
Site de coupure	identique ou proche du site de reconnaissance	24-26 pb en aval du site de reconnaissance	non spécifique et à plus de 1 000 pb du site de reconnaissance
Restriction et méthylation	réactions séparées	simultanées	mutuellement exclusives
ATP nécessaire à la restriction ?	non	oui	oui

- Une séquence hémiméthylée n'est pas reconnue par une enzyme de restriction, mais peut être convertie en une séquence totalement méthylée grâce à une méthylase.

Dans la bactérie, la méthylation sert à perpétuer l'état de modification. La réplication d'un ADN totalement méthylé produit de l'ADN hémiméthylé (voir Figure 15.1). La reconnaissance des séquences hémiméthylées est probablement le mode normal d'action des méthylases *in vivo*.

- Une séquence cible non méthylée peut, *in vitro*, être le substrat d'une réaction de modification ou de restriction. Dans la cellule, l'ADN non modifié est plus enclin à subir une restriction.

La majorité des enzymes de restriction de type II coupe l'ADN aux séquences non méthylées. Une liaison phosphodiester est coupée sur chacun des brins. Cette coupure se produit séquentiellement. Certaines enzymes coupent les brins de façon décalée, d'autres engendrent des extrémités nettes. Une sous-classe d'enzymes coupe l'ADN quelques bases plus loin que le site de reconnaissance.

La méthylase ajoute seulement un groupement méthyl à la fois. Si le substrat n'est pas méthylé, l'enzyme introduit un groupement méthyl puis se dissocie de l'ADN. Le second groupement méthyl ne sera ajouté que s'il se produit un second événement de reconnaissance du site par une méthylase.

LES ACTIVITÉS ALTERNATIVES DES ENZYMES DE TYPE I

La seconde catégorie d'activité de restriction et de modification regroupe des enzymes de type I et de type III qui sont des multimères porteurs à la fois des fonctions endonucléasiques et des fonctions de méthylation. Leur mode d'action est légèrement différent aussi bien des enzymes de type II, que les unes des autres. La proportion de bactéries possédant un système de type I ou de type III n'est pas connue, mais elle est inférieure à celle des bactéries possédant un système de type II.

Les enzymes de **type I** représentées par les activités EcoK et EcoB des souches K et B d'*E. coli*, ont été les premières à être découvertes. Une enzyme de modification et de restriction de type I consiste en trois types de sous-unités, comme l'indique la **Figure 15.2**.

La sous-unité R est responsable de la restriction, la sous-unité M de la méthylation et la sous-unité S de la reconnaissance de la séquence cible sur l'ADN. Lorsque la séquence cible a été reconnue par la sous-unité S, la liaison de l'enzyme est suivie soit de la restriction, soit de la modification. Les activités des sous-unités R et M sont mutuellement exclusives.

Chaque sous-unité est codée par un gène unique. Les mutants dans *hsdR* sont phénotypiquement *r⁻m⁺*; ils ne peuvent pas couper l'ADN mais sont capables de le modifier. Les mutants dans *hsdS* sont *r⁻m⁻*; ils ne

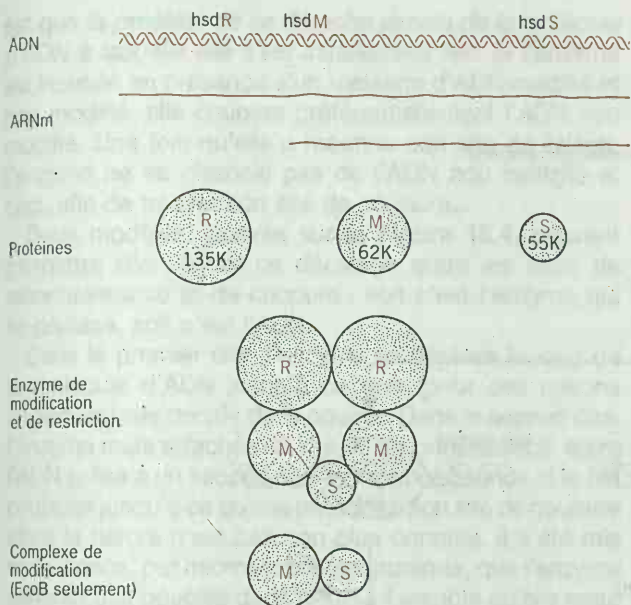


Figure 15.2

Différentes sous-unités sont responsables de la restriction, de la modification et de la reconnaissance chez les enzymes multifonctionnelles de type I.

L'enzyme EcoK a un poids moléculaire d'environ 400 000 daltons. Elle comporte deux copies de la sous-unité R de 135 000 daltons, deux copies de la sous-unité M de 62 000 daltons, et une copie de la sous-unité S de 55 000 daltons.

L'enzyme EcoB est constituée de deux sous-unités de taille identique mais qui peuvent être présentes à différentes concentrations molaires. Les sous-unités M et S de la souche B peuvent former un complexe 1 : 1 qui exerce une fonction de méthylation indépendante de la fonction de restriction. Ce doit être un phénomène naturel car les trois gènes font partie d'un opéron et sont situés les uns par rapport aux autres de la façon suivante : *P1-hsdR-P2-hsdM-hsdS* où *P1* et *P2* sont des promoteurs indépendants. Par conséquent, *hsdM* et *hsdS* peuvent être exprimés séparément de *hsdR*.

peuvent ni couper, ni modifier l'ADN. Les mutants dans *hsdM* sont incapables de restriction et de modification. La sous-unité M est donc impliquée dans la fonction de la sous-unité R. Ce mécanisme est un système de sécurité afin de s'assurer que les mutants *hsdM* ont bien un phénotype *r⁻m⁻* plutôt que *r⁺m⁻* qui serait probablement létal (car l'activité de restriction dégraderait l'ADN non modifié de la cellule).

Les enzymes EcoB et EcoK sont alléliques. Leurs séquences cibles sont différentes, mais dans les cellules diploïdes, la sous-unité S peut diriger les activités des sous-unités R et M d'autres souches bactériennes. Cette

observation confirme le fait que l'étape de reconnaissance est indépendante des événements ultérieurs de restriction ou de méthylation.

Les sites de reconnaissance d'EcoB et d'EcoK sont des structures bipartites consistant en une séquence spécifique de 3 pb séparée de quelques paires de bases d'une autre séquence spécifique de 4 pb. La séparation des deux parties de ce site de reconnaissance signifie que l'une et l'autre sont localisées sur une même face de l'ADN. La séquence de la région intermédiaire ne semble pas importante.

Aucun des deux sites de reconnaissance n'est symétrique, mais chacun d'entre eux possède une adénine méthylée sur un brin différent, comme l'indiquent les astérisques situées sur la séquence de reconnaissance de EcoB.

```

      *
T G A N N N N N N N N T G C T
A C T N N N N N N N N A C G A
      *

```

La détermination d'un ADN donné à être coupé ou modifié est donnée par l'état de la séquence cible, comme l'indique la Figure 15.3. Si la séquence cible est totalement méthylée, l'enzyme s'y lie éventuellement mais se détache aussitôt sans avoir agi. Si la séquence cible est hémiméthylée, l'enzyme méthyle le brin non méthylé afin de perpétuer l'état de méthylation de l'ADN. Si la séquence cible n'est pas du tout méthylée, sa reconnaissance par l'enzyme déclenche la réaction de restriction.

Les enzymes de type I font preuve d'une grande originalité dans la reconnaissance des sites de reconnaissance et de coupure. *En effet, le site de coupure est localisé à plus de 1 000 pb du site de reconnaissance.* Bien que la coupure ne se produise pas à une séquence spécifique, le choix du site de coupure ne se fait pas totalement au hasard car certaines régions de l'ADN sont préférentiellement coupées.

(Ce décalage entre site de reconnaissance et site de coupure nous empêche de déterminer les sites de reconnaissance en fonction des coupures dans l'ADN ; les séquences cibles sont identifiées grâce à leur état de méthylation et par la localisation de mutations supprimant la reconnaissance.)

La réaction de coupure se fait en deux temps. Un seul des brins d'ADN est coupé, suivi peu de temps après par l'autre brin. Il peut y avoir une dégradation exonucléolytique dans les régions entourant le site de coupure. Il se produit une importante hydrolyse d'ATP dont la fonction n'est pas connue.

Comment l'enzyme peut-elle reconnaître un site et en cliver un autre si éloigné ? Une caractéristique importante

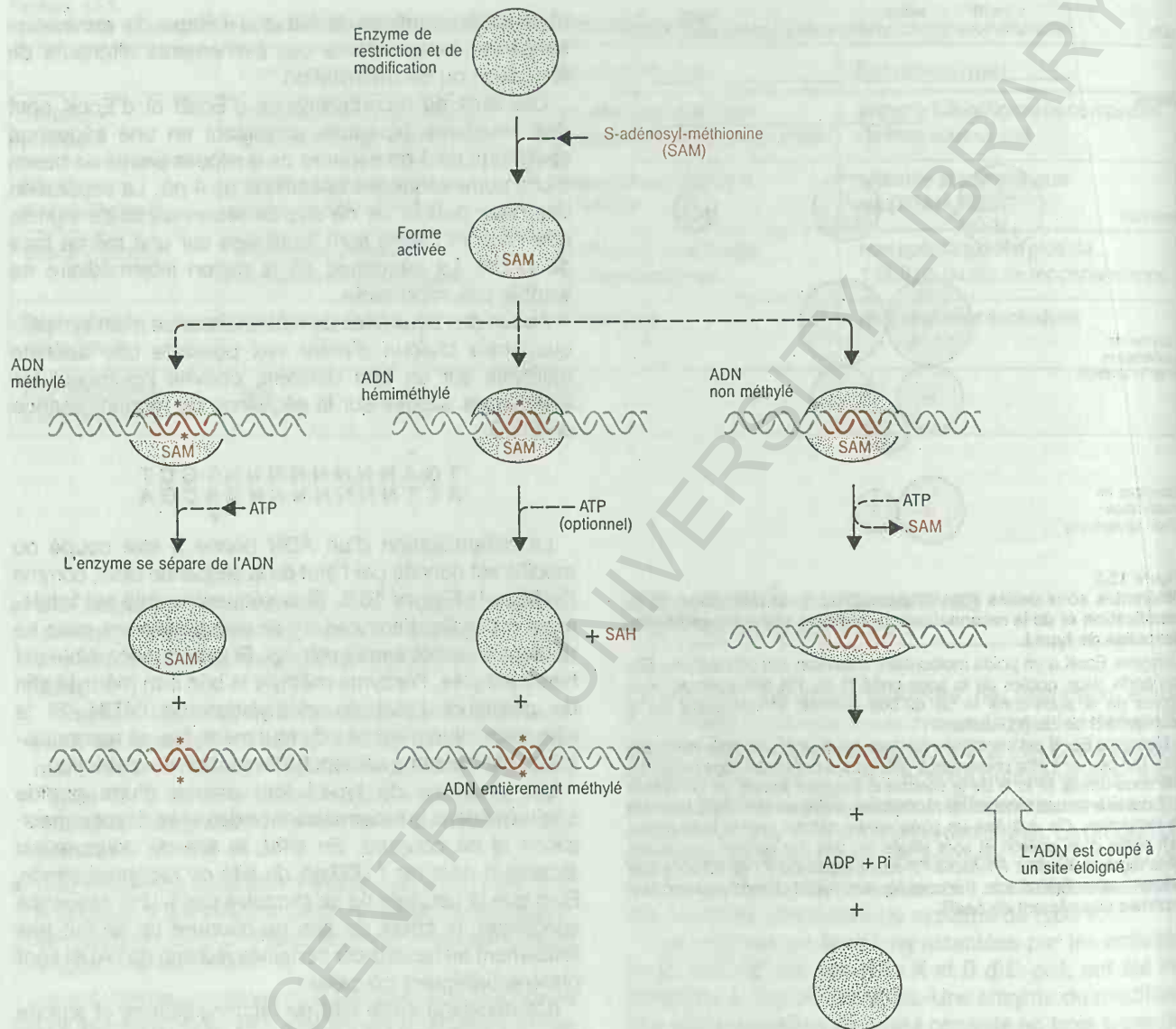


Figure 15.3

Les enzymes de type I se lient à l'ADN au niveau de sites cibles. Ils se séparent de l'ADN totalement méthylé, complètent la méthylation de l'ADN aux sites hémiméthylés et se déplacent le long de l'ADN à partir des sites non méthylés afin de couper la molécule un peu plus loin.

Les groupements nécessaires à la méthylation sont apportés par le cofacteur S-adenosyl-méthionine (SAM) qui est converti en S-adenosyl-homocystéine (SAH) après la réaction. La SAM se lie à la sous-unité M. Dans la première étape de la réaction, la SAM est un effecteur allostérique qui permet à la sous-unité S de changer de conformation afin qu'elle puisse se lier à l'ADN.

L'étape suivante est une réaction avec l'ATP. Si l'enzyme est liée à un site entièrement méthylé, l'ATP lui permet de se séparer de l'ADN. A un site non méthylé, l'ATP rend l'enzyme capable de restriction. Cette transformation touche la sous-unité R. L'hydrolyse d'ATP est nécessaire dans la réaction de clivage. La SAM est libérée de l'enzyme avant que les étapes de restriction ne commencent.

est que la protéine ne se détache jamais de la molécule d'ADN à laquelle elle s'est initialement liée. Si l'enzyme est incubée en présence d'un mélange d'ADN modifié et non modifié, elle coupera préférentiellement l'ADN non modifié. Une fois qu'elle a reconnu son site de liaison, l'enzyme ne se dissocie pas de l'ADN non méthylé et ceci, afin de trouver son site de coupure.

Deux modèles, illustrés sur la **Figure 15.4**, peuvent permettre d'expliquer ce décalage entre les sites de reconnaissance et de coupure : soit c'est l'enzyme qui se déplace, soit c'est l'ADN.

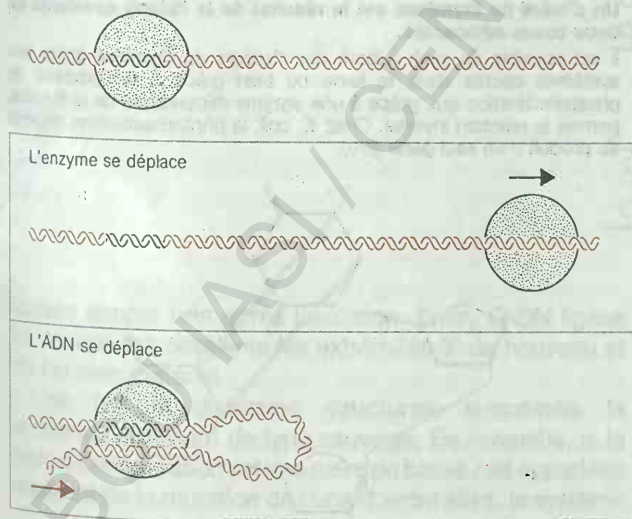
Dans le premier cas, l'enzyme se déplace le long de la molécule d'ADN jusqu'à ce que (pour des raisons inconnues) elle décide de le couper. Dans le second cas, l'enzyme reste attachée au site de reconnaissance, attire l'ADN grâce à un second site de reconnaissance et le fait coulisser jusqu'à ce qu'elle rencontre son site de coupure (dont la nature n'est pas non plus connue). Il a été mis en évidence, par microscopie électronique, que l'enzyme génère des boucles dans l'ADN ; il semble qu'elle reste attachée au site de reconnaissance après la coupure, ce qui est en faveur du second modèle.

LA DOUBLE ACTIVITÉ DES ENZYMES DE TYPE III

Trois enzymes de restriction et de modification de type III ont été étudiées ; EcoP1 et EcoP15 sont codées par les plasmides P1 et P15 d'*E. coli*, Hinf est trouvé dans *H.*

Figure 15.4

Une enzyme de type I se déplace-t-elle le long de l'ADN ou bien reste-t-elle fixée à son site cible tout en attirant l'ADN afin qu'il traverse la protéine ?



influenzae de sérotype R₁. Chacune de ces enzymes est composée de deux types de sous-unités. La sous-unité R est responsable de la restriction, la sous-unité MS de la modification et de la reconnaissance.

Les activités de restriction et de modification sont exprimées *simultanément*. Tout d'abord, l'enzyme se fixe à l'ADN sur son site de liaison ; cette réaction est dépendante d'ATP (la dépendance d'ATP pour les étapes ultérieures varie selon les enzymes). Puis les activités de restriction et de modification entrent *en compétition pour réagir avec l'ADN*.

La méthylation se produit au site de liaison, ce qui est en accord avec le fait que la sous-unité MS regroupe les activités de méthylation et de reconnaissance. La coupure de restriction se produit 24-26 bases plus loin, probablement parce que l'enzyme est suffisamment large pour que la sous-unité de restriction puisse prendre contact avec l'ADN à cet endroit, comme le décrit schématiquement la **Figure 15.5**. La restriction engendre des coupures échelonnées, distantes de 2-4 bases.

Les enzymes méthylent les résidus adénine, mais les sites cibles de P1 et P15 ont une caractéristique surprenante : ils ne sont méthylés que sur un seul brin. Les séquences de ces sites sont les suivantes :

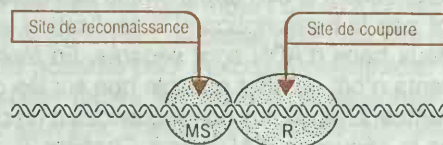
sP1	AGACC	sP15	CAGCAG
	TCTGG		GTCTGC

Comment l'état de méthylation se perpétue-t-il ? La **Figure 15.6** montre que la réplication engendre deux types de sites. Une des répliques a le brin méthylé initial ; en fait, elle est identique à la molécule parentale. L'autre réplique n'est pas du tout méthylée. Elle devient donc, en principe, la cible soit de la restriction, soit de la modification. Qu'est-ce qui favorise la méthylation plutôt que la restriction ? Nous ne connaissons pas la réponse, mais il semble que la réaction de modification soit liée à la réplication.

Figure 15.5

Les enzymes de type III sont constituées de deux sous-unités : la sous-unité MS qui assure la reconnaissance et la méthylation du site cible, et la sous-unité R responsable de la restriction qui se produit près du lieu où elle se situe.

La sous-unité R est de 106 000-110 000 daltons ; le polypeptide MS est de 73 000-80 000 daltons.



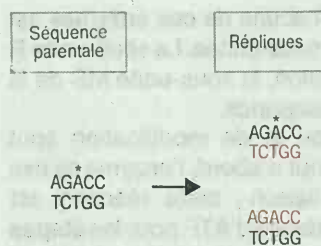


Figure 15.6

La réplication d'une séquence sP1 méthylée donne lieu à une réplique méthylée et une réplique non méthylée.

FACE À DES DÉGATS DE L'ADN !

Les dommages subis par l'ADN sont limités grâce à des systèmes qui s'aperçoivent de l'apparition de changements et les rectifient. Les systèmes de réparation sont sans nul doute aussi complexes que l'appareil de réplication lui-même, ce qui permet de mesurer leur importance pour la survie de la cellule. Le taux de mutation est le reflet de la balance qui existe entre le nombre de dommages subis par l'ADN et le nombre de dommages corrigés (ou non corrigés).

On appelle « dommage » tout changement aboutissant à dévier l'ADN de sa structure en double hélice. Entrent dans cette définition les événements suivants :

- L'introduction de cassures simple brin.
- La suppression d'une base laissant la base complémentaire non appariée.
- La liaison covalente de nouveaux groupes à une base ayant pour conséquence la modification de ses propriétés.
- La conversion d'une base en une autre qui n'est pas correctement appariée à la base complémentaire.
- L'introduction de liaisons covalentes entre bases de la même chaîne d'ADN ou de chaînes différentes.

Les systèmes de réparation reconnaissent certaines distorsions dans l'ADN qui représentent en quelque sorte leurs signaux d'action. Les événements cités précédemment peuvent, en fait, être divisés en deux catégories.

Les **mutations ponctuelles** peuvent être reconnues car elles empêchent le bon appariement des bases. Elles n'affectent pas la transcription ni la réplication, une fois que les deux brins d'ADN sont séparés. En réalité, ces changements n'ont un effet nuisible que sur les générations suivantes, effet dû aux conséquences d'un changement dans la séquence d'ADN (voir Chapitre 2).

Les **distorsions structurales** peuvent être un obstacle physique à la réplication ou à la transcription. Une cassure ou l'absence d'une base sur un brin d'ADN peut l'empêcher de servir de matrice pour la synthèse d'ADN ou d'ARN. Les structures engendrées par les liaisons covalentes anormales à l'intérieur d'un même brin ou entre deux brins opposés peuvent inhiber la réplication et la transcription. L'exemple de distorsion structurale le plus étudié est celui, engendré par irradiation aux ultraviolets, d'une liaison covalente entre deux thymines adjacentes produisant un **dimère de thymines**, dessiné sur la Figure 15.7.

Une cellule peut avoir plusieurs systèmes lui permettant de rectifier les aberrations de l'ADN. Il en existe deux types généraux. La partie endommagée est excisée et remplacée grâce à un **système de réparation**. Si le dommage subi persiste au site initial, les chaînes filles engendrées après la réplication seront réparées grâce à un **système de rétablissement**.

Les systèmes d'**excision-resynthèse** ôtent les bases mal appariées ou endommagées de l'ADN, puis synthétisent une nouvelle région d'ADN afin de remplacer les bases manquantes. Les différentes étapes de ce système sont illustrées sur la Figure 15.8.

Pour l'étape d'**incision**, une endonucléase coupe le brin d'ADN des deux côtés de la structure endommagée.

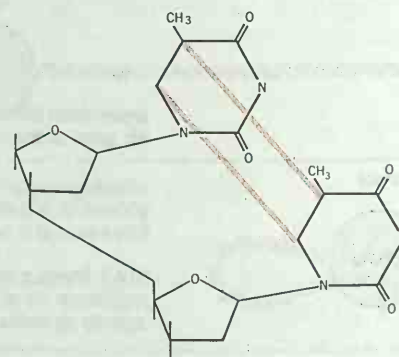
Pour l'étape d'**excision**, une exonucléase 5'-3' enlève le segment d'ADN indésirable.

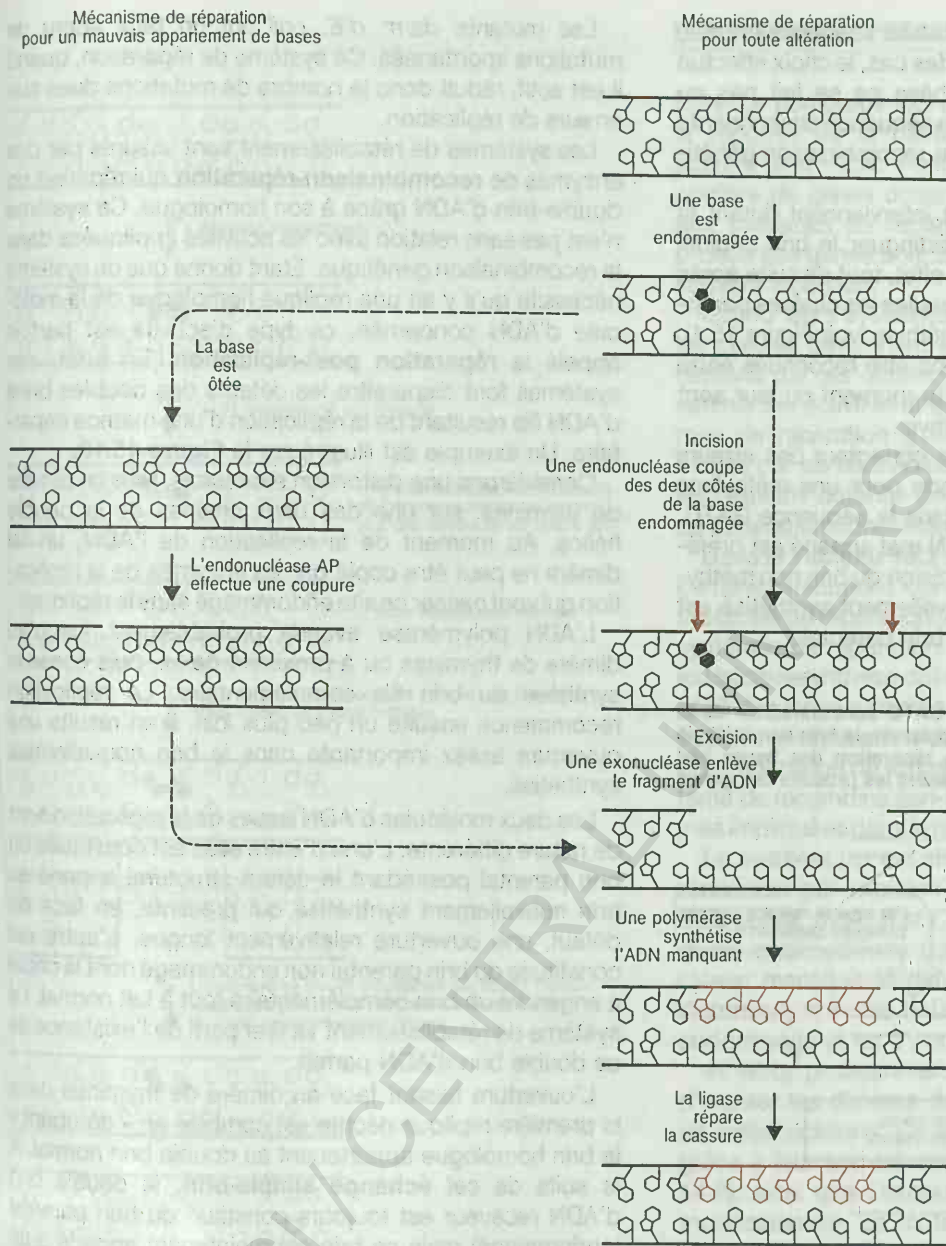
Pour l'étape de **resynthèse**, une ADN polymérase synthétise la séquence manquante en copiant la région

Figure 15.7

Un dimère de thymines est le résultat de la liaison covalente de deux bases adjacentes.

Il est possible de venir à bout des dimères de thymines grâce aux systèmes décrits dans le texte ou bien grâce à un procédé de photoréactivation qui, grâce à une enzyme dépendante de la lumière, permet la réaction inverse. Chez *E. coli*, la photoréactivation dépend du produit d'un seul gène (*phr*).



**Figure 15.8**

Le système d'excision-resynthèse permet d'enlever et de remplacer un segment d'ADN comportant la (les) base(s) endommagée(s) mais aussi les bases avoisinantes.

A droite : Le mécanisme général de la réparation nécessite la coupure des deux côtés de la base endommagée suivie de l'excision.

A gauche : Le mécanisme permettant de supprimer l'uracile de l'ADN enlève seulement la base concernée, après quoi une endonucléase coupe du côté 5' de la base conduisant à une excision.

laissée simple brin après l'excision. Enfin, l'ADN ligase lie de manière covalente les extrémités 3' du nouveau et de l'ancien matériel.

Une fois la distorsion structurale supprimée, la séquence redevient de type sauvage. En revanche, si la cible de la réparation est une paire de bases mal appariées résultant de la mutation de l'une d'entre elles, le système

de réparation n'a pas de moyen intrinsèque pour savoir quelle est celle des deux qui est la base du type sauvage et celle qui l'est du type muté ! Ce qu'il constate est un mauvais appariement de bases dont chacune représente une cible possible pour l'excision-resynthèse.

Si la base mutée est excisée, la séquence redevient de type sauvage. Mais si c'est la base originelle (de type

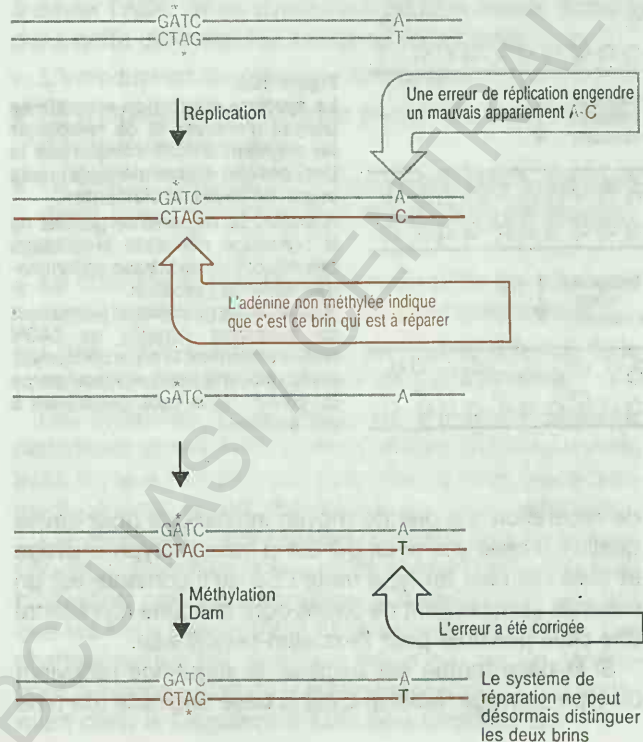
sauvage) qui est excisée, la nouvelle séquence (mutée) devient stable. Dans la majorité des cas, le choix effectué par le système excision-resynthèse ne se fait pas au hasard, mais il est influencé par la séquence environnante (qui peut être importante dans la recombinaison génétique ; voir Chapitre 28).

Si des erreurs d'appariement interviennent durant la réplication, il est possible de distinguer le brin original d'ADN de celui qui est muté. En effet, tout de suite après la réplication de l'ADN méthylé, seules les chaînes parentales portent les groupements méthyls (voir Figure 15.1). Les molécules filles peuvent donc être reconnues entre le moment de leur synthèse et le moment où leur sont introduits des groupements méthyls.

Ceci nous amène au système correcteur des erreurs de réplication. Le gène *dam* code pour une méthylase dont la cible est une adénine dans la séquence GATC. La Figure 15.9 montre que l'ADN mal apparié est préférentiellement réparé grâce à l'excision du brin non méthylé. Par conséquent, le brin nouvellement synthétisé est corrigé grâce à la séquence du brin parental.

Figure 15.9

Après la réplication, les séquences GATC sont les cibles de la méthylase Dam. Avant qu'il y ait méthylation, le brin non méthylé constitue la cible des systèmes de réparation des bases mal appariées grâce à un système comprenant les produits des gènes *mutH*, *mutL*, *mutS* et *uvrE*.



Les mutants *dam⁻* d'*E. coli* ont un taux accru de mutations spontanées. Ce système de réparation, quand il est actif, réduit donc le nombre de mutations dues aux erreurs de réplication.

Les systèmes de rétablissement sont assurés par des enzymes de **recombinaison-réparation** qui réparent un double-brin d'ADN grâce à son homologue. Ce système n'est pas sans relation avec les activités impliquées dans la recombinaison génétique. Étant donné que ce système nécessite qu'il y ait une réplique homologue de la molécule d'ADN concernée, ce type d'activité est parfois appelé la **réparation post-réplication**. En effet, ces systèmes font disparaître les défauts des doubles brins d'ADN fils résultant de la réplication d'une matrice imparfaite. Un exemple est illustré sur la Figure 15.10.

Considérons une distorsion structurale, telle un dimère de thymines, sur une des deux chaînes de la double hélice. Au moment de la réplication de l'ADN, un tel dimère ne peut être copié par les enzymes de la réplication qui vont passer ce site endommagé sans le répliquer.

L'ADN polymérase avance probablement jusqu'au dimère de thymines ou à proximité de lui, puis cesse la synthèse du brin fils complémentaire. La réplication recommence ensuite un peu plus loin. Il en résulte une ouverture assez importante dans le brin nouvellement synthétisé.

Les deux molécules d'ADN issues de la réplication sont de nature différente. L'une d'entre elles est constituée du brin parental possédant le défaut structural apparié au brin nouvellement synthétisé qui présente, en face du défaut, une ouverture relativement longue. L'autre est constituée du brin parental non endommagé dont la copie a engendré un brin complémentaire tout à fait normal. Le système de rétablissement va tirer parti de l'existence de ce double brin d'ADN parfait.

L'ouverture faisant face au dimère de thymines dans la première réplique décrite est comblée en « dérobant » le brin homologue appartenant au double brin normal. À la suite de cet **échange simple-brin**, le double brin d'ADN receveur est toujours constitué du brin parental (endommagé) mais ce brin est maintenant apparié à un brin de type sauvage. En revanche, le double brin d'ADN donneur est représenté par un brin parental intact apparié à un brin possédant maintenant une ouverture ; cette ouverture est ensuite comblée par la synthèse de réparation afin d'engendrer une molécule tout à fait normale. Ainsi, le défaut initial reste confiné à ce qu'il était. En réalité, le système de recombinaison-réparation est utilisé après chaque cycle de réplication à moins ou jusqu'à ce que le défaut soit éliminé par le système d'excision-resynthèse.

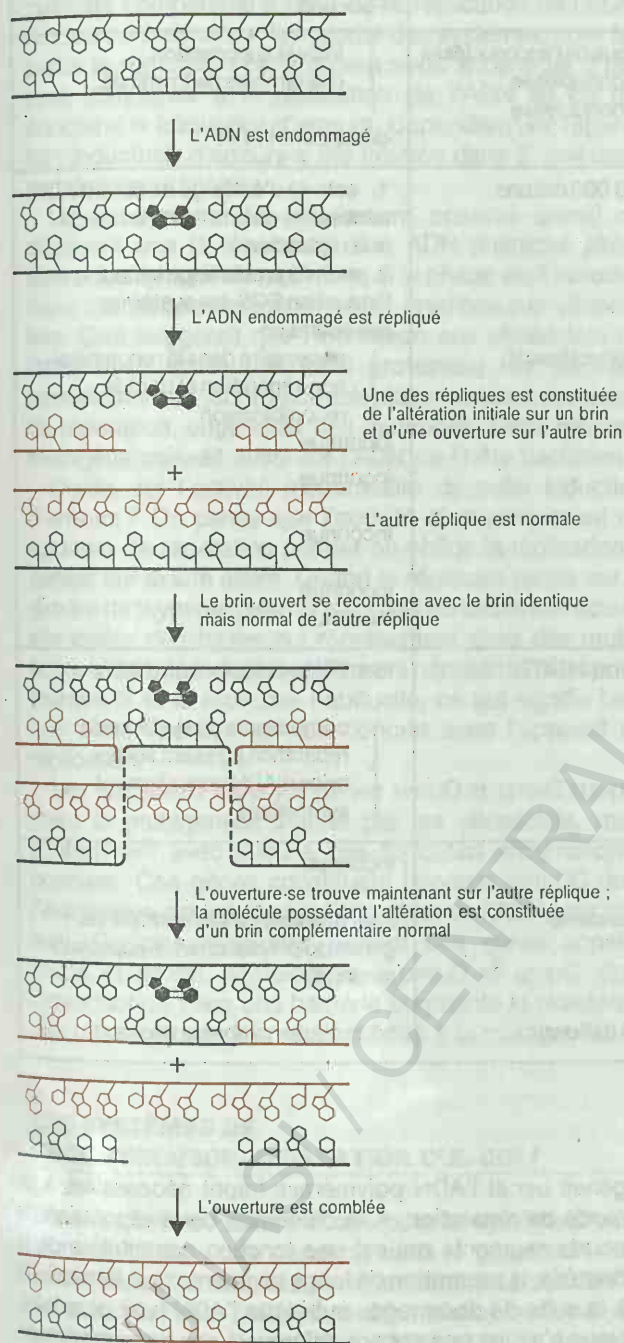
LES SYSTÈMES D'EXCISION-RESYNTHESE D'*E. COLI*

Figure 15.10

La recombinaison-réparation utilise un brin normal d'ADN pour combler une ouverture située sur le brin nouvellement synthétisé en face d'une altération non réparée.

Dans de nombreux loci d'*E. coli*, les mutations affectent la réponse de la bactérie aux agents extérieurs qui altèrent son ADN. Font partie de ces loci, un grand nombre de gènes codant pour les enzymes participant aux systèmes de réparation de l'ADN ; cependant, le produit des gènes concernés ainsi que leur fonction sont souvent inconnus.

Quand leurs systèmes de réparation sont inexistants, les cellules deviennent *extrêmement* sensibles aux irradiations par les ultraviolets. Le test le plus utilisé pour estimer les activités et les efficacités relatives des systèmes de réparation a été l'introduction de dommages induits par les ultraviolets. Il faut garder en mémoire que les résultats obtenus peuvent être différents en fonction de l'altération considérée.

Dans le **Tableau 15.2** sont résumées les propriétés de certaines mutations qui empêchent *E. coli* d'utiliser ses systèmes de réparation. Nous connaissons au moins trois systèmes de réparation. Deux modes de réparation par excision-resynthèse ont été identifiés grâce à l'hétérogénéité des longueurs de segments d'ADN réparés. Ils sont appelés la **réparation de courts fragments**, d'une part, et la **réparation de longs fragments**, d'autre part. Le système de recombinaison-réparation utilise certaines enzymes impliquées dans la recombinaison (voir Chapitre 28).

Le système *uvr* est impliqué dans les deux modes de réparation par excision-resynthèse ; il est constitué de trois gènes, *uvrA*, *B* et *C* qui codent pour les sous-unités d'une endonucléase de réparation. *In vitro*, l'endonucléase reconnaît le dimère de thymine (ou une autre distorsion) et incise l'ADN à deux endroits, en 5' du site endommagé et en 3', environ 12 nucléotides plus loin.

In vitro, plusieurs enzymes d'*E. coli* sont capables d'exciser les dimères de thymine de l'ADN ayant subi de telles incisions. Ce sont les ADN polymérases I et III grâce à leur activité exonucléasique 5'-3' (voir Tableau 14.2), ainsi que l'exonucléase VII spécifique de l'ADN monocaténaire. La mutation de l'une de ces nucléases ne diminue pas de façon notable la capacité d'*E. coli* à éliminer les dimères de thymine. Il est probable que l'ADN polymérase I soit responsable de la plupart des excisions *in vivo*. L'activité hélicasique de *uvrD* est requise lors de la réaction.

La longueur moyenne du segment d'ADN excisé est d'environ 20 nucléotides, ce qui a suscité, pour ce système de réparation, l'appellation de réparation de courts segments. L'enzyme impliquée dans la synthèse de réparation est probablement l'ADN polymérase I (les ADN polymérases II et III peuvent toutefois la remplacer).

Tableau 15.2

De nombreux produits de gènes sont impliqués dans les systèmes de réparation de l'ADN chez *E. coli*.

Gène	Conséquence de mutation <i>in vivo</i>	Produit du gène	Fonction
<i>uvrA</i> <i>uvrB</i> <i>uvrC</i> <i>uvrD</i>	sensibilité aux UV » » »	sous-unité ATPasique de l'endonucléase sous-unité de l'endonucléase sous-unité de l'endonucléase ADN hélicase II	} initie la suppression des dimères de thymine désapparie l'ADN
<i>recA</i>	1. déficience dans la recombinaison 2. non induction des systèmes de réparation après altérations de l'ADN	Protéine Rec de 40 000 daltons	1. activité d'échange de brins d'ADN nécessaire dans la recombinaison et la réparation 2. activité protéasique initiant l'induction SOS des systèmes de réparation
<i>recB</i> <i>recC</i>	déficience dans la recombinaison »	sous-unité de l'exonucléase V »	} nécessaire dans la recombinaison-réparation et dans la recombinaison
<i>sbcB</i> <i>recF</i>	supprime la mutation <i>recBC</i> } déficience dans la recombinaison-réparation et dans la	exonucléase I inconnu	inconnue inconnue
<i>recJ</i>	} recombinaison des mutants } <i>recBCsbcB</i>	inconnu	inconnue
<i>recK</i> <i>recE</i>	déficience dans la recombinaison	inconnu exonucléase VIII	inconnue inconnue
<i>dam</i>	sensibilité aux UV, taux de mutation accru	méthylase agissant sur GATC	identifie le « vieux » brin d'ADN
<i>mutH</i> <i>mutL</i> <i>mutS</i> <i>uvrE</i>	» » » »	inconnu » » »	} composants du système de réparation agissant sur les brins d'ADN nouvellement synthétisés
<i>umuC</i> <i>lon</i>	supprime les effets mutagènes de la réparation inductrice d'erreurs sensibilité aux UV, déficience du septum	inconnu protéase ATP dépendante qui se lie à l'ADN inconnu	inconnue il est possible qu'elle contrôle les gènes du polysaccharide capsulaire inconnue
<i>mut</i> <i>lexA</i>	> 6 gènes dont la mutation augmente le taux de mutation supprime la réponse SOS	répresseur (22 000 daltons)	contrôle de nombreux gènes

Dans le système *uvr*, d'autres gènes dont la mutation empêche la réparation de courts segments ont été identifiés, mais leur produit ou fonction n'est pas connu.

La réparation de courts segments est responsable de 99 % des événements d'excision-resynthèse. Le 1 % restant est assuré par le remplacement de segments d'ADN longs d'environ 1 500 nucléotides en général, mais pouvant aller jusqu'à plus de 9 000 nucléotides. Les

gènes *uvr* et l'ADN polymérase I sont nécessaires à ce mode de réparation. A la différence de la réparation de courts segments qui est une fonction constitutive de la bactérie, la réparation de longs segments doit être induite à la suite de dommages subis par l'ADN (voir plus loin). Nous n'avons pas encore différencié ces deux modes de réparation en terme de produits de gènes impliqués dans l'un ou l'autre des systèmes.

Les systèmes de réparation qui conduisent à la synthèse d'ADN doivent posséder un système de contrôle qualitatif comparable à celui de la réplication de l'ADN. A notre connaissance, la majorité des systèmes, dont fait partie la réparation par excision sous le contrôle d'*uvr*, sont identiques à la réplication de l'ADN en ce qui concerne la fréquence d'erreurs. Cependant une réparation inductrice d'erreurs a été trouvée dans *E. coli* dans certaines conditions.

Un tel système fut initialement observé quand on découvrit que la réparation d'un ADN phagique altéré était accompagné de mutations si le phage était introduit dans des cellules précédemment irradiées aux ultraviolets. Cela suggérait que l'irradiation aux ultraviolets de l'hôte avait activé une (des) protéine(s) non active(s) normalement et dont l'activité, dans certains systèmes de réparation, engendrait des mutations. Cette réponse mutagène agissait aussi sur l'ADN de l'hôte bactérien.

Quelle est l'activité responsable de cette induction d'erreurs ? On pense que l'activité d'un composant du système de réparation permet ou oblige la réplication à passer sur le site altéré. Quand la réplicase passe sur le dimère de thymine, elle n'insère pas les bases correctes ; elle insère des bases qui représentent alors des mutations. L'activité inductrice d'erreurs nécessite l'ADN polymérase III et la réplicase habituelle, ce qui signifie bien que cette fonction agit de concert avec l'appareil de réplication.

Les mutations dans les gènes *umuD* et *umuC* suppriment la mutagenèse induite par les ultraviolets, mais n'interfèrent avec aucune des fonctions enzymatiques connues. Ces gènes constituent l'opéron *umuDC* dont l'expression est induite par l'altération de l'ADN (voir plus loin). Certains plasmides portent des gènes appelés *mucA* et *mucB*, homologues à *umuD* et *umuC*, dont l'introduction dans une bactérie augmente la résistance aux ultraviolets et à la susceptibilité à la mutagenèse.

LES SYSTÈMES DE RECOMBINAISON-RÉPARATION D'E. COLI

Dans les cellules défectueuses pour le système d'excision-resynthèse, une mutation dans le gène *recA* supprime toutes les activités de réparation restantes. Si l'on tente de répliquer l'ADN dans des cellules *uvr-recA⁻*, il y a production de fragments d'ADN dont la longueur correspond à la distance attendue entre deux dimères de thymine. Ce résultat suggère que si les dimères ne

sont pas éliminés, ils font obstacle à la réplication en l'absence de la fonction *RecA* et sont léthaux pour la cellule. On comprend mieux qu'un double mutant ne puisse tolérer plus d'1-2 dimères dans son génome (alors qu'une bactérie de type sauvage peut en supporter jusqu'à 50).

Le gène *recA* et des mutations dans d'autres gènes *rec* ont permis l'identification des systèmes de recombinaison-réparation. Les éléments impliqués dans ces systèmes sont communs à ceux de la recombinaison mais ne sont cependant pas tout à fait identiques car certaines mutations n'affectent que l'un des deux systèmes.

Les mutants *recA* sont presque totalement défectueux pour la recombinaison génétique et en systèmes de rétablissement. La fonction de la protéine *RecA* est l'échange de brins entre molécules d'ADN ; cette fonction représente l'activité essentielle de la recombinaison (voir Chapitre 28) et est impliquée dans l'échange simple-brin de la recombinaison-réparation.

Les propriétés des doubles mutants suggèrent que la protéine *recA* participe à deux systèmes *Rec*. Afin de déterminer si deux gènes font partie du même ou de deux systèmes différents, les phénotypes des simples mutants sont comparés à ceux des doubles mutants. Si les gènes appartiennent au même système, le double mutant sera dénué de deux activités au lieu d'une et son phénotype sera d'autant plus sévère. Selon ce critère, ont été déterminés deux systèmes, l'un impliquant les gènes *recBC*, l'autre le gène *recF*.

Les gènes *recBC* codent pour les deux sous-unités de l'exonucléase V dont l'action est contrôlée par d'autres composants du système (chez les mutants *recA*, l'enzyme n'est pas contrôlée et dégrade une quantité massive d'ADN. C'est le phénotype « reckless » (imprudent) qui est à l'origine du nom de ces loci). La fonction de *recF* est inconnue. D'autres loci *rec* ont été identifiés grâce à l'effet de certaines mutations sur la recombinaison-réparation et la recombinaison. D'autres encore sont déterminés grâce à leur capacité à supprimer des mutations *rec*. Comme pour les loci *uvr* supplémentaires, leur produit et la fonction de leur produit n'ont pas été découverts.

La dénomination de ces gènes est basée sur les phénotypes des mutants ; cependant, il est important de noter qu'une mutation isolée dans certaines conditions et cataloguée comme un locus *uvr* s'avère être aussi un locus *rec* quand elle est isolée dans d'autres conditions. Nous ne sommes pas capables actuellement de définir le nombre de fonctions qui appartiennent à un système ou de savoir comment les systèmes interagissent entre

eux. Les systèmes *uvr* et *rec* ne sont pas entièrement indépendants l'un de l'autre car les mutants *uvr* ont une efficacité réduite dans la recombinaison-réparation.

Nous devons nous attendre à trouver un réseau d'enzymes constitué de nucléase, polymérase et autres enzymes impliquées dans la réparation dont la fonction peut intervenir dans différents systèmes de réparation (ou une enzyme habituellement utilisée dans tel système peut être substituée par une autre qui normalement intervient dans un autre système).

La participation active de la protéine RecA dans la recombinaison-réparation ne représente qu'un de ses rôles. Cette protéine extraordinaire a une autre fonction très différente. C'est une protéine qui est activée après irradiation aux ultraviolets (ou tout autre traitement qui bloque la réplication). L'activité protéasique est responsable de l'induction de nombreux gènes dont ceux responsables de certaines fonctions de réparation (par exemple, les gènes responsables de la réparation de longs segments).

A cause de la double fonction de la protéine RecA, il est difficile de savoir si un déficit de la recombinaison est dû à la perte de la fonction « d'échange simple-brin d'ADN » de la protéine ou à toute autre fonction dont l'induction dépend de l'activité protéasique de cette même protéine.

UN SYSTÈME SOS CONSTITUÉ DE NOMBREUX GÈNES

De nombreux traitements qui endommagent l'ADN ou inhibent la réplication chez *E. coli* induisent une série de changements phénotypiques appelés la **réponse SOS**. Elle est la conséquence de l'interaction de la protéine RecA avec le répresseur LexA.

L'altération peut être due à l'irradiation par les ultraviolets (le cas le plus étudié) ou à des agents intercalants ou alkylants. L'inhibition de la réplication par l'un de ces moyens ou encore par privation de thymines, adjonction de drogues ou mutations dans les gènes *dna*, a le même effet sur la cellule.

En effet, dans tous les cas, la réponse de la cellule est d'augmenter sa capacité à réparer l'ADN endommagé, et ce en induisant la synthèse des composants du système de réparation de longs segments et du système de recombinaison-réparation Rec. De plus, la division cellulaire est inhibée. Les prophages lysogéniques peuvent être activés.

Le premier événement de cette réponse est l'activation, par le traitement endommageant l'ADN, de l'activité protéasique de la protéine RecA. La nature du signal de cette activation est inconnue. Étant donné que la réponse

SOS est induite par de nombreuses altérations, les travaux actuels s'orientent plutôt vers la recherche d'un intermédiaire commun du métabolisme de l'ADN qui stimulerait la protéine RecA.

Ce signal peut être une petite molécule relarguée de l'ADN, ou une structure adoptée par l'ADN. *In vitro*, l'activité protéasique de RecA ne fonctionne qu'en présence d'une région d'ADN monocaténaire et de l'ATP. Par conséquent, le signal pourrait être la présence d'une région monocaténaire au niveau du site endommagé.

Quel que soit le signal, son interaction avec la protéine RecA est rapide ; la réponse SOS est induite quelques minutes après le traitement mutagène. Une fois activée, la protéase RecA coupe une protéine de 22 000 daltons codée par le gène *lexA*. La protéine LexA est relativement stable dans les cellules sauvages, où elle fonctionne comme répresseur de nombreux opérons. Le clivage protéolytique de ce répresseur induit simultanément tous les opérons. Ce système est illustré dans la **Figure 15.11**.

Parmi les gènes réprimés par LexA (dont un petit nombre est connu) figurent des gènes impliqués dans les fonctions de réparation. Un dépistage systématique des gènes inductibles par LexA est réalisé en fusionnant au hasard des opérons avec le gène *lacZ* et en dosant la quantité de β -galactosidase produite quand les cellules sont traitées par de la mitomycine C. Au moins cinq gènes ont ainsi été identifiés ; ils sont appelés les gènes *din* (« damage inductible » : inductible par l'endommagement de l'ADN). Les gènes connus et faisant partie de la réponse SOS par activation de leur produit, sont *recA*, *lexA*, *uvrA*, *uvrB*, *umuC* et *hima*.

Certains des gènes SOS sont uniquement actifs quand les cellules sont traitées par un mutagène ; d'autres sont actifs dans les cellules sauvages mais leur niveau d'expression est augmenté après le clivage de LexA. Le gène *uvrB*, qui est un composant du système d'excision-résynthèse, est composé de deux promoteurs ; l'un fonctionne indépendamment de LexA, l'autre est sous son contrôle. Par conséquent, après le clivage de LexA, le gène peut être exprimé à partir du second promoteur comme il l'est à partir du premier.

LexA réprime ses gènes cibles en se liant à une région de 20pb d'ADN appelée la **boîte SOS** ; cette séquence est symétrique et une copie en est présente à chaque locus cible. Les boîtes SOS de différents loci ne sont pas identiques mais sont conformes à une séquence consensus constituée de 8 positions toujours conservées. De même que d'autres opérateurs, les boîtes SOS chevauchent leur promoteur respectif. Le locus *lexA*, qui est sujet à une auto-répression, comporte deux boîtes SOS adjacentes.

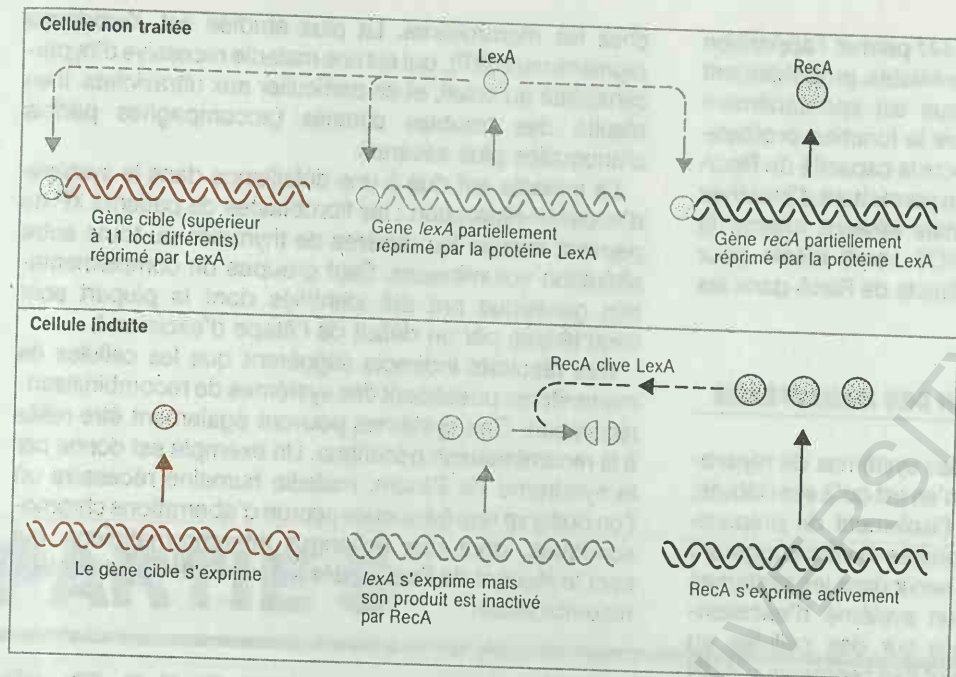


Figure 15.11

La protéine LexA réprime de nombreux gènes, dont font partie *recA*, *lexA* et d'autres gènes impliqués dans la réparation. Quand la protéase RecA est activée, elle clive LexA et induit les gènes normalement réprimés par LexA.

RecA et LexA se prennent mutuellement pour cible dans le circuit SOS : RecA clive LexA qui réprime *recA*. La réponse SOS a pour résultat l'amplification de la protéine RecA et du répresseur LexA. Ceci n'est pas aussi contradictoire qu'il y paraît.

En effet, l'expression accrue de la protéine RecA est nécessaire (probablement) pour assurer le rôle qu'elle joue dans les systèmes de recombinaison-réparation. Sous induction, le taux de RecA passe d'environ 1 200 molécules/cellule à 50 fois plus. Dans les cellules mutées, il y a donc suffisamment de protéase pour que toute protéine LexA supplémentaire soit clivée et éviter une nouvelle répression par LexA.

L'importance majeure de ce système est qu'il permet à la cellule de retourner rapidement à son état initial. Quand le signal inducteur est supprimé, la protéine RecA perd son activité protéasique. Le gène *lexA* a alors un niveau d'expression élevé ; en l'absence de la protéase RecA, il y a accumulation de la protéine LexA sous sa forme entière, ce qui permet l'extinction des gènes SOS. Cela explique sans doute le fait que la réponse SOS soit facilement réversible.

La protéase RecA agit également sur d'autres répresseurs dont ceux de certains prophages. Parmi eux, se trouve le répresseur de lambda (grâce auquel l'activité protéasique fut découverte). C'est la raison pour laquelle le phage lambda est activé après irradiation aux ultraviolets ; en effet, RecA clive le répresseur lysogénique, permettant ainsi au phage d'entrer dans un cycle lytique.

Cette réaction n'est pas une réponse SOS de la cellule mais elle est le signal pour le prophage que la cellule est en difficulté et que sa survie est en danger. Il est donc plus sûr pour le prophage d'entrer dans un cycle lytique et d'engendrer une descendance phagique. Le prophage imite le système cellulaire en répondant au même signal (activation de RecA).

La protéase RecA coupe toutes ses protéines cibles à une séquence dipeptidique Ala-Gly au milieu de la chaîne polypeptidique. Il y a seulement une faible homologie entre les acides aminés situés de part et d'autre de ce dipeptide, ce qui suggère que la structure tertiaire de la protéine est une caractéristique importante de la reconnaissance.

Les deux activités de la protéine RecA sont relativement

indépendantes. La mutation *recA441* permet l'apparition du système SOS sans traitement préalable, probablement parce que la fonction protéasique est spontanément active. Une autre mutation inactive la fonction protéasique. Aucune des mutations n'affecte la capacité de RecA de réagir avec l'ADN. Une mutation permettant d'inactiver la fonction de recombinaison mais laissant intacte la capacité d'induire le système SOS serait idéale pour distinguer les effets directs et indirects de RecA dans les systèmes de réparation.

LES SYSTÈMES DE RÉPARATION DES MAMMIFÈRES

La caractérisation biochimique des systèmes de réparation dans les cellules eucaryotes n'en est qu'à ses débuts. Elle consiste, pour l'instant, en l'isolement de préparations enzymatiques dont les propriétés *in vitro* suggèrent qu'elles peuvent éventuellement servir dans les systèmes de réparation. La présence d'un système d'excision-resynthèse peut être déterminée sur des cellules en culture en suivant le remplacement des segments d'ADN qui ont été altérés par traitement mutagène. Une des techniques consiste à suivre les événements dans des cellules synchrones (en dehors de la phase S) au moment où toute **synthèse d'ADN non programmée** peut être attribuée aux systèmes de réparation.

Les cellules de mammifères montrent une grande hétérogénéité en ce qui concerne le taux d'ADN incorporé après altération. Cependant, les plus longs segments d'ADN synthétisés lors de la réparation dans les cellules de mammifères sont de taille comparable à ceux du système de réparation de courts segments bactérien. Ce système n'agit que sur des altérations causées par l'irradiation aux ultraviolets ou tout traitement ayant un effet identique. Un autre système permet l'introduction de 3-4 bases aux sites endommagés par irradiation aux rayons X ou alkylation.

Certaines maladies héréditaires démontrent l'existence et prouvent l'importance des systèmes de réparation

chez les mammifères. La plus étudiée est *Xeroderma pigmentosum* (XP), qui est une maladie récessive d'hypermotilité au soleil, et en particulier aux ultraviolets. Il en résulte des troubles cutanés (accompagnés parfois d'anomalies plus sévères).

La maladie est due à une défaillance dans le système d'excision-réparation ; les fibroblastes de patients XP ne peuvent exciser les dimères de thymine ou toute autre altération volumineuse. Sept groupes de complémentarité génétique ont été identifiés dont la plupart sont caractérisés par un déficit de l'étape d'excision*.

Des résultats indirects suggèrent que les cellules de mammifères possèdent des systèmes de recombinaison-réparation. Ces systèmes peuvent également être reliés à la recombinaison génétique. Un exemple est donné par le syndrome de Bloom, maladie humaine récessive où l'on observe une fréquence accrue d'aberrations chromosomiques, dont des échanges interchromatidiens, qui sont le résultat de l'action désordonnée des systèmes de recombinaison.

POUR EN SAVOIR PLUS

Les activités indépendantes de restriction et de modification (type II) ont été examinées par **Modrich** (*Quart. Rev. Biophys.* 3, 315-369, 1979). **Yuan** a réalisé une précieuse revue générale concernant les enzymes multifonctionnelles (type I et type III), en insistant plus particulièrement sur les activités enzymatiques (*Ann. Rev. Biochem.*, 50, 285-315, 1981). Il développa un modèle de translocation des enzymes de type I (*Cell* 20, 237-244, 1980). Les fonctions des endonucléases AP et des systèmes qui leur sont associés ont été étudiées par **Lindahl** (*Prog. Nucleic Acid Res.* 22, 135-192, 1979). **Hanawalt et al.** présentèrent une revue générale concernant les systèmes et les mécanismes de réparation chez les procaryotes et les eucaryotes (*Ann. Rev. Biochem.* 48, 783-836, 1979). Le système de contrôle SOS d'*E. coli* a été examiné de façon complète par **Little & Mount** (*Cell* 29, 11-22, 1982) et **Walter** (*Microbiol. Rev.* 48, 60-93, 1984).

* **N. d. T. :** Des travaux analogues sont activement menés sur les autres maladies humaines comportant probablement des anomalies de la réparation. Dans l'anémie de Fanconi, il a été possible de compléter l'anomalie de réparation par transfert d'ADN et d'isoler l'ADN responsable de cette activité de complémentarité. Dans une autre affection, le syndrome de Bloom, un déficit en ADN ligase a été caractérisé par plusieurs équipes.

PARTIE 5

CONSTITUTION DU GÉNOME DES EUCARYOTES

Le concept génétique actuel définit l'aspect d'un organisme comme étant le résultat de l'interaction de tous ses gènes avec l'environnement dans lequel il évolue. Une modification dans l'un de ses gènes, ce que l'on appelle une mutation ponctuelle, est capable de bouleverser l'équilibre du système et est souvent responsable de l'apparition d'une particularité chez cet organisme, généralement une anomalie rendant l'individu inférieur au type sauvage... L'étude de déficiences, c'est-à-dire l'étude d'organismes chez lesquels un ou plusieurs gènes sont manquants, montre que la majorité des déficiences sont léthales pour l'organisme quand elles sont présentes à l'état homozygote. Cela signifie que la présence d'une grande majorité de gènes est essentielle pour permettre la vie de l'organisme. Des travaux effectués sur des déficiences chez *D. melanogaster* indiquent qu'elles sont nombreuses à être léthales pour la cellule, et prouvent que même un petit groupe de cellules possédant une déficience à l'état homozygote mais situé dans des tissus normaux ne peut survivre. Cela suggère que des gènes sont actifs dans chaque cellule et que probablement, la majorité d'entre eux joue un rôle extrêmement important dans les processus vitaux pour la cellule.

Milislav Demerec, 1935

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

LES MÉTHODES DE RÉGULATION DES IMPACTS

La régulation des impacts des projets de développement est un processus complexe. Elle implique la mise en œuvre de diverses méthodes de régulation. Ces méthodes sont souvent classées en deux catégories : les méthodes de régulation préventive et les méthodes de régulation corrective.

Les méthodes de régulation préventive visent à éviter ou à réduire les impacts négatifs des projets de développement avant qu'ils ne se produisent. Elles comprennent des mesures telles que l'évaluation d'impact, la consultation des parties prenantes, et la mise en place de mesures de protection.

Les méthodes de régulation corrective visent à corriger les impacts négatifs des projets de développement après qu'ils se sont produits. Elles comprennent des mesures telles que la mise en place de mesures de réparation, la mise en place de mesures de compensation, et la mise en place de mesures de réhabilitation.

La régulation des impacts des projets de développement est un processus continu. Elle nécessite une surveillance constante des impacts des projets de développement et une mise à jour des mesures de régulation en fonction des résultats de la surveillance.

La régulation des impacts des projets de développement est un processus complexe. Elle implique la mise en œuvre de diverses méthodes de régulation. Ces méthodes sont souvent classées en deux catégories : les méthodes de régulation préventive et les méthodes de régulation corrective.

La régulation des impacts des projets de développement est un processus continu. Elle nécessite une surveillance constante des impacts des projets de développement et une mise à jour des mesures de régulation en fonction des résultats de la surveillance.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

Les auteurs ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît. Ils ont pu constater que les impacts des projets de développement sont souvent plus complexes qu'il n'y paraît.

PARTIE 2

CONSTITUTION DU GÉNOME DES EUKARYOTES

BCU IASI / CENTRAL UNIVERSITY LIBRARY

CHAPITRE 16

L'EXTRAORDINAIRE PUISSANCE DE LA TECHNOLOGIE DES RECOMBINANTS D'ADN

La technologie permettant de manipuler l'ADN est si importante qu'est devenu chose courante de construire de nouvelles molécules d'ADN en liant des séquences issues de sources différentes. Le produit qui en résulte est souvent appelé un **recombinant d'ADN**, et les techniques utilisées (plus familièrement) le **génie génétique**. Ces techniques sont utilisables aussi bien chez les procaryotes que chez les eucaryotes, même si elles semblent plus particulièrement intéressantes en ce qui concerne les génomes eucaryotes.

Dans ce chapitre, nous allons nous intéresser aux premières étapes de la caractérisation des gènes. Comment isoler l'ARN (à partir du cytoplasme) ou l'ADN (du génome) d'un gène particulier ? Une fois ses acides nucléiques identifiés, comment obtenir suffisamment de matériel pour sa caractérisation ? Peut-on diriger une séquence codante que l'on a isolée, pour la synthèse de son produit *in vitro* ou *in vivo* ? Peut-on introduire des changements dans cette séquence qui vont influencer son expression et nous permettre de découvrir la nature des signaux qui la régulent ?

Le **clonage** d'un fragment d'ADN permet la production, à partir d'une seule molécule, de quantités indéfinies de cette molécule. (Un **clone** est défini comme étant un grand nombre de cellules ou molécules identiques toutes issues d'une même cellule ou molécule ancestrale). Le clonage d'ADN est rendu possible grâce à la capacité des plasmides bactériens et des phages à conserver leurs propriétés après introduction dans leur génome de séquences d'ADN supplémentaires.

Une telle insertion engendre un plasmide ou phage **hybride** ou **chimérique**, constitué d'une part de l'ADN de son génome originel, et d'autre part de séquences supplémentaires « étrangères ». Ces éléments chimériques se répliquent dans les bactéries tout comme le plasmide ou phage dont ils sont issus ; ils peuvent donc être obtenus en grande quantité. Des copies du fragment étranger inséré peuvent être recouvrées à partir de la descendance. Les propriétés des espèces chimériques n'étant habituellement pas modifiées par la présence des séquences étrangères, pratiquement n'importe quelle séquence d'ADN peut être clonée de cette manière. Le

phage ou plasmide étant utilisé comme véhicule pour l'ADN étranger, qui est une partie inerte du génome, ils sont souvent appelés **vecteurs de clonage**.

TOUTE SÉQUENCE D'ADN PEUT ÊTRE CLONÉE DANS UNE BACTÉRIE

Les molécules d'ADN hybrides sont construites grâce à la propriété des enzymes de restriction de couper l'ADN en des séquences particulières et relativement courtes de nucléotides (voir Chapitre 4). En coupant le vecteur et l'ADN à cloner en des sites appropriés, il est possible de construire des molécules hybrides afin d'amplifier la quantité de matériel et ainsi permettre l'expression d'une séquence particulière.

Une caractéristique essentielle pour un vecteur de clonage est de posséder un site où l'ADN étranger pourra être inséré sans pour cela supprimer une de ses fonctions vitales. La méthode la plus simple est d'utiliser une enzyme de restriction qui n'a qu'une seule séquence cible située à un endroit non essentiel dans l'ADN du vecteur. La réaction d'insertion de l'ADN étranger dans le vecteur n'engendre qu'une faible proportion de génomes chimériques ; il est donc important de posséder des moyens de *sélection* qui permettront de discerner le génome chimérique du vecteur initial.

Les génomes plasmidiques sont toujours circulaires. Par conséquent, une coupure unique transforme l'ADN en une molécule linéaire comme l'illustre la **Figure 16.1**. Les deux extrémités du vecteur peuvent se lier aux extrémités de l'ADN étranger linéaire, et régénérer ainsi un plasmide chimérique circulaire. Il n'y a pas de limite formelle en ce qui concerne la taille de l'ADN étranger pouvant être inséré. Le plasmide chimérique peut se perpétuer indéfiniment dans la bactérie. Il peut être isolé grâce à sa taille et grâce au fait qu'il est circulaire, par électrophorèse sur gel par exemple.

De nombreux plasmides portent des gènes de résistance aux antibiotiques. Cette caractéristique est très utile dans la construction du système de clonage. Habituellement, on utilise un plasmide possédant les gènes de résistance à deux antibiotiques. Un des gènes est utilisé pour identifier la bactérie qui porte le plasmide, l'autre pour distinguer les plasmides chimériques des vecteurs parentaux. Si le site d'insertion de l'ADN étranger est situé dans le second gène, le plasmide chimérique perd sa résistance à l'antibiotique. Par conséquent, le vecteur parental conserve la résistance aux deux antibiotiques alors que le plasmide chimérique n'est plus résistant qu'à l'un des antibiotiques et est devenu sensible à l'autre.

Les vecteurs de clonage ont été construits en amélio-

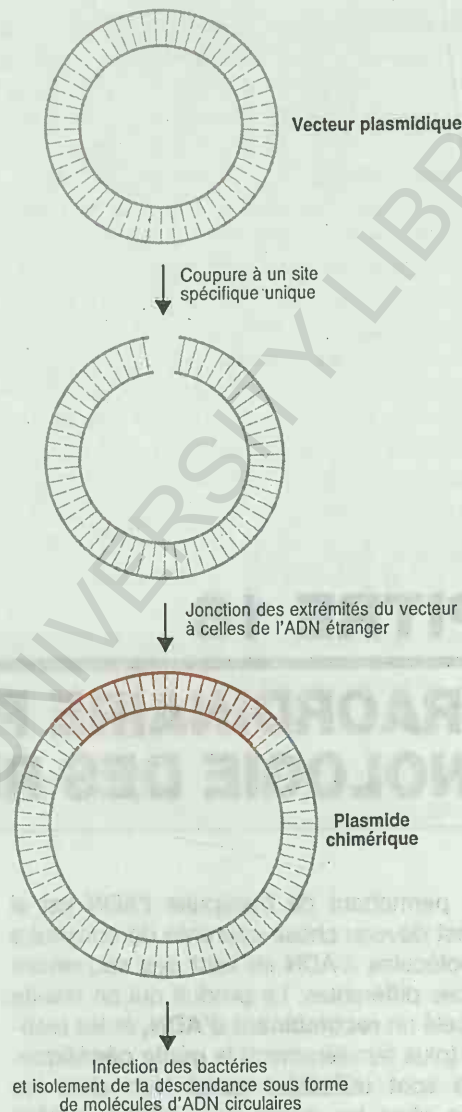


Figure 16.1
L'ADN étranger peut être cloné dans des vecteurs plasmidiques par insertion en un site approprié.

rant les plasmides naturels de sorte qu'ils possèdent toutes les propriétés désirées pour un clonage aisé. Cette manipulation peut conduire à la modification du système contrôlant la réplication ou à l'addition de gènes conférant la résistance à des antibiotiques particuliers. Un des vecteurs de clonage couramment utilisé est **pBR322**, dérivant de vecteurs de clonage construits précédemment et ayant subi diverses altérations séquentielles. Ce plasmide existe en plusieurs copies dans la bactérie et porte la résistance à la tétracycline et à l'ampicilline ; il possède, situés à des endroits judicieux, des sites uniques de coupure pour plusieurs enzymes de restriction.

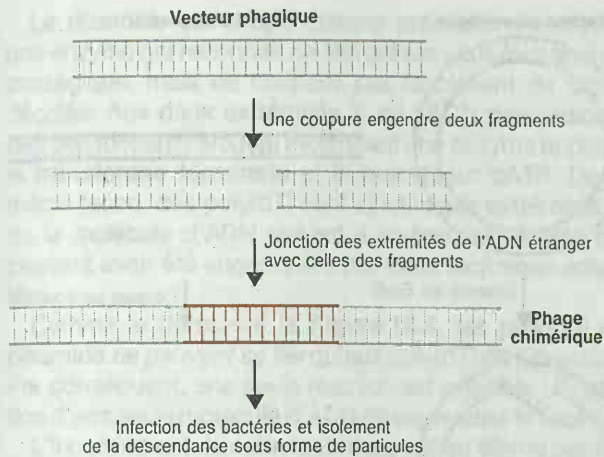


Figure 16.2

L'ADN étranger peut être cloné dans des vecteurs phagiques par insertion dans une région non essentielle du génome.

Les phages apportent un autre type de système de vecteur. Le phage est habituellement constitué d'une molécule d'ADN linéaire. Par conséquent, la coupure de son ADN à un site de restriction unique engendre deux fragments. Ces deux fragments peuvent être liés à l'ADN étranger afin d'engendrer un phage chimérique comme l'illustre la **Figure 16.2**. Il est facile d'isoler le phage chimérique si on le laisse entrer dans le cycle lytique et produire de multiples particules. Cette technique est toutefois limitante en ce qui concerne la taille de l'ADN étranger à insérer. En effet, la tête du phage a la capacité d'empêcher l'empaquetage de tout ADN trop long dans les particules de la descendance.

Afin de remédier à ce problème, une partie de l'ADN du vecteur ne contenant pas de gènes essentiels pour le phage peut être *remplacée* par de l'ADN étranger. Ce système a été grandement exploité en ce qui concerne le phage lambda ; en effet, à partir de ce phage, un nouveau vecteur a été créé en manipulant l'ADN afin de diminuer au maximum la taille du génome (en déléant les gènes non essentiels) et pour qu'il ne possède qu'un seul site de restriction EcoRI. Le vecteur de clonage ainsi déléé, est en fait *trop court pour être empaqueté dans la tête du phage*, qui a également une norme quant à la taille minimum d'ADN à empaqueter. Il est donc impératif qu'un fragment d'ADN étranger soit inséré dans le vecteur parental précédemment coupé, pour que soit engendré

un phage qui puisse se multiplier. Cette condition équivaut à un système automatique de sélection pour obtenir des phages chimériques.

Le développement des systèmes permettant l'empaquetage « *in vitro* » de l'ADN dans des particules de phages a rendu ce type de vecteur encore plus intéressant. Afin de combiner les propriétés intéressantes des plasmides d'une part et des phages d'autre part, on a construit des **cosmides**. Ce sont des plasmides dans lesquels ont été insérées des séquences d'ADN particulières (les sites cos) qui sont nécessaires à l'empaquetage de l'ADN de lambda dans les particules. Ces vecteurs peuvent toujours se perpétuer dans les bactéries sous la forme d'un plasmide ; ils peuvent également être purifiés par empaquetage *in vitro* dans des phages. Ils sont toujours sujets à une limitation de taille imposée par la tête du phage, mais une quantité plus importante d'ADN étranger peut être empaquetée car les gènes du phage sont inutiles et donc absents*.

Nous avons considéré les vecteurs de clonage en pensant utiliser uniquement des hôtes bactériens. Il est cependant parfois préférable d'utiliser des hôtes eucaryotes. Les seuls plasmides eucaryotes connus résident dans la levure**. Par construction de recombinants d'ADN, des plasmides « à double emploi » ou plus simplement des plasmides « navette » ayant les séquences nécessaires pour survivre et dans *E. coli* et dans *S. cerevisiae*, ont été obtenus. Ces vecteurs peuvent donc être utilisés dans les deux systèmes.

CONSTRUIRE UN ADN CHIMÉRIQUE

La liaison d'un fragment d'ADN à un vecteur de clonage nécessite que se produise une réaction entre les extrémités du fragment et celles du vecteur. Le fragment et le vecteur peuvent se recombiner pour donner une molécule d'ADN chimérique grâce à la production de séquences complémentaires à leurs extrémités.

Généralement, les enzymes de restriction qui effectuent des **coupures décalées** sont utilisées afin d'obtenir de courtes extrémités monocaténaïres et complémentaires, que l'on appelle extrémités à bouts **cohésifs**. L'enzyme EcoRI est un exemple classique, elle coupe chacun des deux brins de la molécule d'ADN à un endroit différent se situant de chaque côté d'une séquence palindromique qui fait partie du site de reconnaissance de l'enzyme.

* N. d. T. : De nouveaux vecteurs de clonage sont continuellement développés : les améliorations concernent les marqueurs de sélection, la multiplicité des sites de clonage (présence de « polylinker »), la facilité de déterminer la séquence nucléotidique des fragments introduits, voire de les transcrire. Une limitation des programmes de clonage réside en la difficulté de faire pénétrer l'ADN recombinant dans les bactéries. Grâce à leur capsid, les particules phagiques permettent d'introduire de l'ADN de phage ou de cosmide avec une très grande efficacité ; l'introduction d'ADN non empaqueté, par exemple de plasmide, est relativement moins efficace.

** N. d. T. : Des vecteurs de clonage eucaryotiques demeurant extrachromosomiques sont également utilisés dans les cellules animales. Ils sont en général dérivés de génomes viraux : adénovirus, virus de papillome bovin et vaccine.

Dans une séquence d'ADN où *N* représente les bases non impliquées dans la reconnaissance, l'enzyme *EcoRI* coupe dans le site de reconnaissance (représenté en couleur) au niveau des liaisons indiquées par des flèches verticales.

5'...pNpNpNpGpApApTpTpCpNpNpN...3'
3'... NpNpNpCpTpTpApApGpNpNpN...5'

Les fragments d'ADN de chaque côté du site cible se séparent. Ils sont constitués de régions monocaténaïres sortantes complémentaires à leurs extrémités, qui sont dues au décalage qui existe entre les points de coupure. Ces extrémités à bouts cohésifs sont indiquées en couleur.

et

pApApTpTpCpNpNpN...3'
GpNpNpNp...5'

5'...pNpNpNpG
3'... NpNpNpCpTpTpApAp

Les extrémités à bouts cohésifs peuvent s'apparier grâce à leur complémentarité. Quand deux molécules différentes sont coupées par *EcoRI*, les extrémités à bouts cohésifs produites sont *identiques* sur les deux molécules. Elles peuvent donc se lier ensemble comme l'illustre le protocole décrit sur la **Figure 16.3**. Cette réaction effectuée, il y a production d'un plasmide chimérique intact à ceci près qu'il n'y a pas de liaisons covalentes entre le vecteur et l'ADN étranger. *In vitro*, l'ADN ligase se charge de remédier à cette défaillance. La technique de recombinaison de deux molécules d'ADN a des *avantages* mais aussi des *inconvénients*.

Un des avantages est que le plasmide chimérique possède des sites *EcoRI* qui ont été recréés aux deux extrémités de l'ADN inséré. Par conséquent, l'ADN étranger pourra facilement être recouvré à partir des copies du vecteur chimérique en utilisant l'enzyme *EcoRI*.

Un des inconvénients est que les extrémités à bouts cohésifs *EcoRI* du vecteur s'apparient entre elles et reforment le vecteur sans qu'il y ait insertion d'ADN étranger, ou bien que le vecteur insère un long segment d'ADN constitué de plusieurs fragments d'ADN étranger mis bout à bout. Ceci démontre la nécessité de sélectionner des plasmides qui ne sont enrichis que d'une seule insertion.

L'inconvénient de n'employer que des enzymes de restriction pour la production d'extrémités utilisables dans la réaction de jonction, est que leur site de reconnaissance ne se situe pas toujours, dans la séquence d'ADN étrangère, à des endroits stratégiques. Une autre technique permet à n'importe quelle extrémité d'être utilisable dans la recombinaison avec un plasmide.

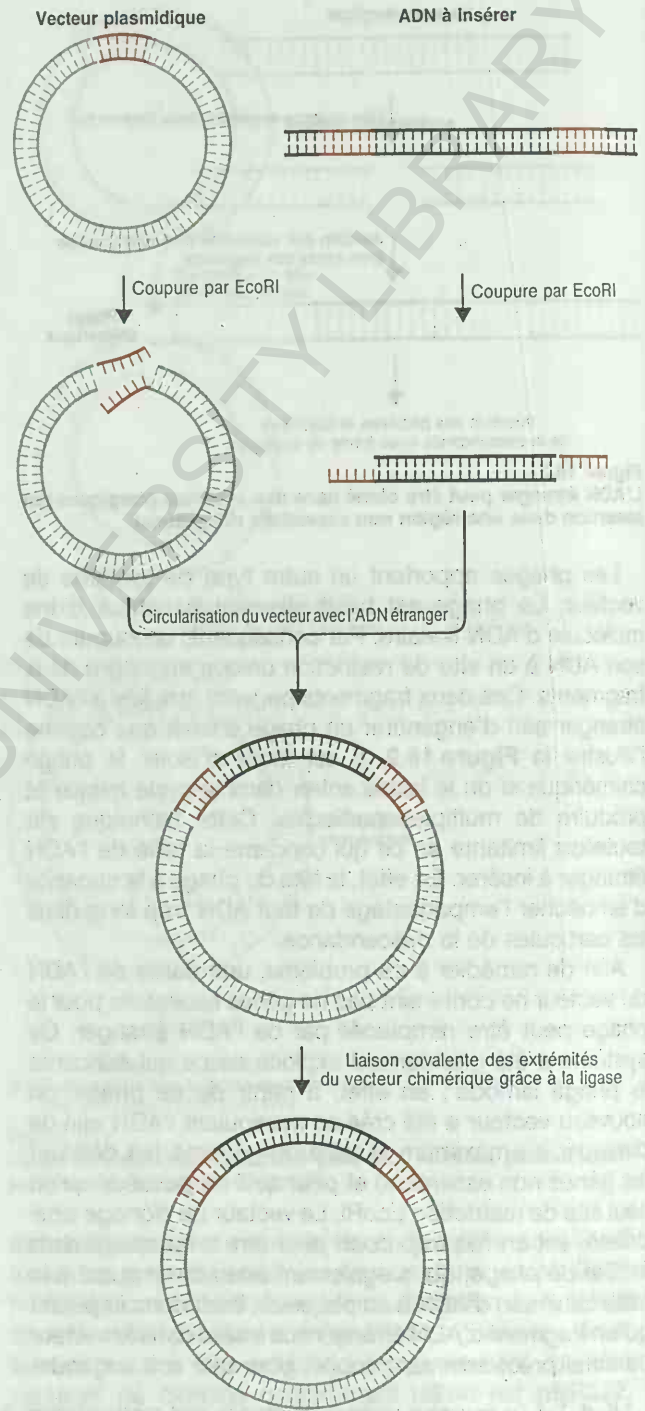


Figure 16.3

Toute séquence d'ADN qui est contenue entre deux sites de coupure *EcoRI* peut être insérée dans un vecteur plasmidique possédant un site *EcoRI*, par coupure puis circularisation des deux molécules ensemble.

Le plasmide est coupé comme précédemment avec une enzyme qui reconnaît un site unique situé à un endroit stratégique, mais ne coupant pas forcément de façon décalée. Aux deux extrémités 3' de l'ADN plasmidique, des poly(dA) sont ajoutés en utilisant une enzyme appelée la **transférase terminale** et le précurseur dATP. De la même façon, des poly(dT) sont ajoutés aux extrémités 3' de la molécule d'ADN qui est à insérer, extrémités qui peuvent avoir été engendrées par toute technique adaptée.

Comme le démontre la **Figure 16.4**, les poly(dA) du plasmide *ne peuvent se lier* qu'aux poly(dT) de l'insertion. Par conséquent, une seule réaction est possible : l'insertion d'une seule molécule d'ADN étranger dans le vecteur.

L'inconvénient de cette technique est qu'elle ne permet pas une excision aisée du fragment inséré à partir du plasmide chimérique cloné, car le site de reconnaissance de l'enzyme de restriction existant dans le vecteur initial a été supprimé par l'insertion d'ADN étranger. La séquence insérée est entourée de chaque côté par des régions d'appariement poly(dA) : poly(dT).

Une autre technique très utile est la **liaison d'extrémités franches**, qui repose sur une des propriétés de l'ADN ligase de T4 qui est de joindre des molécules d'ADN possédant des extrémités franches, c'est-à-dire dépourvues, à leurs extrémités, de séquences monocaténaïres sortantes (cette propriété est à ajouter à l'activité habituelle de restauration des liaisons covalentes cassées dans le double-brin d'ADN). Si l'on utilise des enzymes de restriction qui coupent les deux brins d'ADN au même endroit, on peut avoir directement recours à la technique de liaison d'extrémités franches pour joindre les fragments d'ADN ensemble.

Le grand avantage de cette technique est qu'elle permet la jonction de n'importe quelles extrémités ensemble sans se soucier de leur séquence. Ceci est particulièrement utile quand on veut lier deux séquences définies sans introduire de matériel supplémentaire entre elles. Un des problèmes inhérent à cette technique est que l'on ne contrôle pas la nature des extrémités franches qui s'unissent entre elles. Il est donc nécessaire, après la réaction, d'isoler les produits recherchés parmi tous les produits engendrés*.

Il existe de nombreuses variantes de ces techniques. L'une d'entre elles utilise de petits double-brins d'ADN (« linkers ») contenant la séquence palindromique de reconnaissance de EcoRI (ou toute autre enzyme de restriction). Les « linkers » sont synthétisés chimiquement, et sont liés de façon covalente aux extrémités d'un plasmide ou d'un insert par liaison d'extrémités franches.

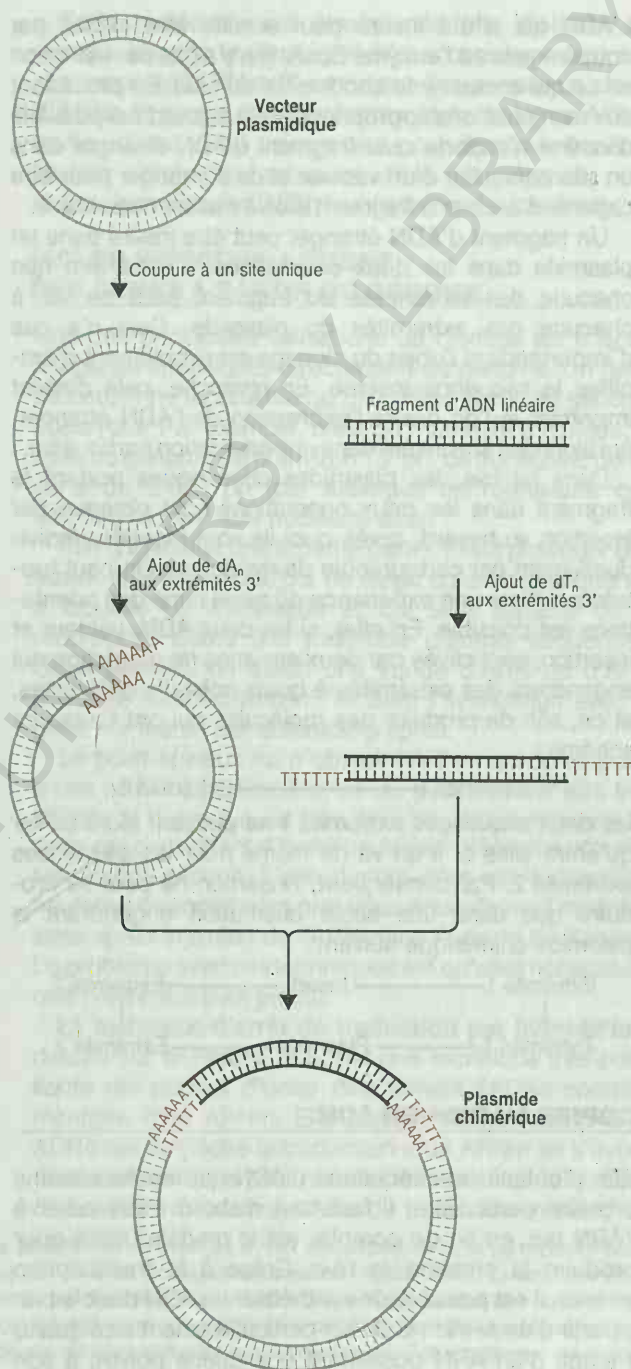


Figure 16.4

La technique d'addition de queues poly (dA : dT) permet la jonction de deux molécules d'ADN, l'une à laquelle ont été ajoutés des poly (dA) aux extrémités 3', et l'autre à laquelle ont été ajoutés des poly (dT) aux extrémités 3' également.

* N. d. T. : Un autre désavantage de la liaison de bouts francs est sa moindre efficacité par rapport à celle d'extrémités cohésives.

L'ADN qui a été inséré peut ensuite être excisé par coupure grâce à l'enzyme *EcoRI* (il n'y a pas de restriction en ce qui concerne le choix initial du site). En procédant aux manipulations appropriées, il est aujourd'hui possible d'insérer n'importe quel fragment d'ADN étranger dans un site particulier d'un vecteur et de s'arranger pour être capable d'exciser le fragment si cela s'avère nécessaire.

Un fragment d'ADN étranger peut être inséré dans un plasmide dans les deux orientations, c'est-à-dire que chacune des extrémités du fragment peut se lier à chacune des extrémités du plasmide. Cela n'a pas d'importance si l'objet du clonage est simplement d'amplifier la séquence insérée. En revanche, cela devient important si l'on désire l'expression de l'ADN étranger, qui alors doit être inséré dans une orientation particulière.

Dans ce cas, les plasmides chimériques portant le fragment dans les deux orientations sont obtenus par insertion au hasard, après quoi ils sont identifiés individuellement par cartographie de restriction. On peut toutefois réaliser une expérience où seule l'une des orientations est possible. En effet, si les deux ADN, vecteur et insertion, sont clivés par deux enzymes de restriction qui engendrent des extrémités à bouts cohésifs différentes, et ce, afin de produire des molécules qui ont toutes ce schéma :

Extrémité 1 ————— Extrémité 2

les deux séquences extrémité 1 ne peuvent alors se lier qu'entre elles et il en va de même pour les séquences extrémité 2. Par conséquent, l'insertion ne peut se produire que dans une seule orientation engendrant le plasmide chimérique suivant :

Extrémité 1	Insert	Extrémité 2	
	Plasmide	Extrémité 2	

COPIER L'ARNm EN ADN

Afin d'obtenir une séquence d'ADN qui représente une protéine particulière, il faut tout d'abord s'intéresser à l'ARN qui, en fin de compte, est le modèle utilisé pour produire la protéine *in vivo*. Grâce à la transcription inverse, il est possible de synthétiser un ADN double brin à partir d'un ARNm. Cela est particulièrement aisé quand il s'agit d'un ARN possédant une queue poly(A) à son extrémité 3', comme l'illustre la Figure 16.5.

Tout d'abord, une **amorce** se lie à la queue poly(dA). Il s'agit d'une courte séquence d'oligo(dT), dont le rôle est de fournir une extrémité 3' qui pourra ensuite être utilisée par une enzyme appelée transcriptase inverse. Cette enzyme permet l'élongation de l'amorce dans la

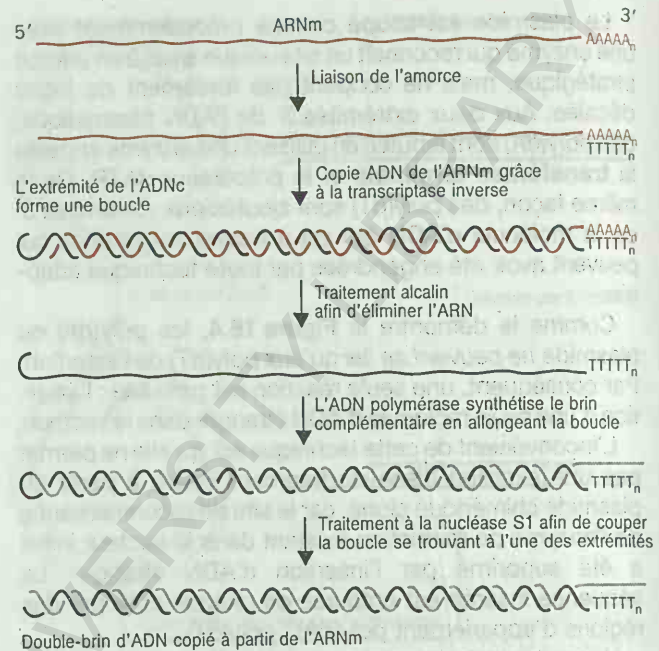


Figure 16.5

L'ARNm peut être copié en une molécule d'ADN bicaténaire.

direction habituelle 5'-3', en ajoutant un par un des déoxynucléotides, dont la détermination est donnée par la complémentarité de l'appariement avec l'ARNm copié.

Le produit de la réaction est une molécule hybride constituée du brin d'ARN initial apparié à un brin complémentaire d'ADN. Le seul problème pratique de cette technique est que la transcriptase inverse a tendance à s'arrêter avant d'avoir atteint l'extrémité 5' de l'ARNm. Dans ce cas, le transcrit inverse obtenu est court et ne représente pas l'ARNm entier car il lui manque les séquences complémentaires de l'extrémité 5'. Cependant, en ajustant judicieusement les conditions expérimentales, il est généralement possible de persuader la transcriptase inverse d'avancer jusqu'au bout de l'ARNm.

Une réaction très utile se produit à l'extrémité de l'ARNm. En effet, l'enzyme « pousse » le transcrit à faire une boucle sur lui-même, en utilisant les quelques dernières bases du transcrit inverse comme modèle pour la synthèse d'un brin complémentaire. Cela signifie que l'extrémité de l'ADN complémentaire à l'ARNm est utilisée pour diriger la synthèse d'une courte séquence qui est identique à l'ARNm, et qui le déplace. Cela crée une boucle en épingle à cheveux, généralement longue de 10-20 pb.

L'ARNm initial est ensuite dégradé par traitement alcalin (ce qui n'affecte pas l'ADN). Le produit de la

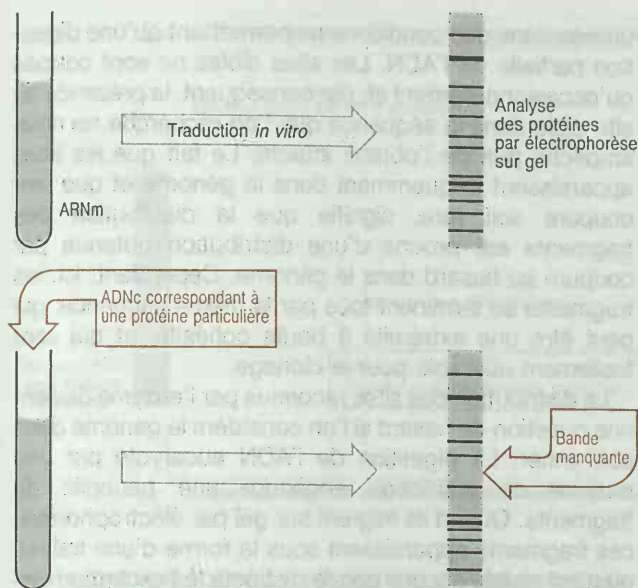


Figure 16.6

La technique d'arrêt de traduction par hybridation permet d'empêcher la synthèse d'une protéine particulière *in vitro*. Les protéines traduites *in vitro* à partir d'une préparation d'ARNm peuvent être séparées par électrophorèse sur gel, et les ADNc peuvent être testés pour leur capacité à empêcher l'apparition d'une bande particulière sur le gel.

réaction est un ADN simple-brin complémentaire de l'ARNm qui est appelé **ADNc**.

La boucle en épingle à cheveux de l'extrémité 3' de l'ADNc fournit l'amorce idéale nécessaire à l'étape suivante, qui consiste à utiliser l'ADN polymérase I d'*E. coli* pour convertir l'ADNc simple-brin en un ADN double-brin grâce à la synthèse du brin complémentaire. Dans cette réaction, l'enzyme se sert de l'ADNc comme matrice pour la synthèse d'une séquence identique à celle de l'ARNm initial. Le produit obtenu est une molécule bicaténaire possédant une boucle en épingle à cheveux à l'une de ses extrémités. Cette boucle est coupée grâce à une enzyme appelée la nucléase S1 (qui dégrade spécifiquement l'ADN simple-brin), ce qui engendre une molécule d'ADN tout à fait normale.

* N. d. T. : Comme toutes les méthodes de recombinaison d'ADN, la constitution de banques d'ADN complémentaires et particulièrement la première étape qui est la synthèse d'ADNc a considérablement évolué et de nouvelles stratégies apparaissent régulièrement. La tendance en ce milieu d'année 1988 est de synthétiser le 2^e brin grâce à l'action combinée de deux enzymes, la ribonucléase H (spécifique des doubles brins hybrides ADN-ARN) et l'ADN polymérase de *E. coli*. En digérant l'ARN encore hybridé au brin complémentaire de l'ADN synthétisé par la transcriptase inverse, la ribonucléase H engendre de multiples amorces ribonucléotidiques qui sont allongées par la polymérase. Cette méthode, d'exécution simple et rapide, permet de ne pas perdre d'informations de séquences à l'extrémité 5' du transcrit, conséquence inéluctable de la digestion de la boucle en épingle à cheveux par la nucléase S1.

** N. d. T. : Là encore, l'évolution des techniques de criblage de banques d'ADNc a été considérable. Outre les techniques décrites dans le texte, deux méthodes sont très utilisées actuellement : l'étude de l'expression du gène cloné et l'hybridation avec des oligonucléotides de synthèse dont la séquence a été déterminée grâce à la connaissance d'une petite portion de la séquence protéique. La première méthode consiste à insérer la séquence à cloner à l'intérieur du gène codant pour une protéine du vecteur. La protéine hybride synthétisée par *E. coli* peut être détectée, le plus souvent par un anticorps, parfois grâce à certaines de ses propriétés fonctionnelles.

Cette molécule peut être clonée afin d'obtenir en grande quantité le gène synthétique correspondant à la séquence de l'ARNm. C'est ce que l'on appelle un **clone d'ADNc** (à cause de cette terminologie, le terme ADNc, au lieu d'être utilisé uniquement pour décrire le transcrit inverse simple-brin, est aussi utilisé par extension pour décrire la molécule d'ADN double-brin qui a été cloné)*.

ISOLER INDIVIDUELLEMENT DES GÈNES A PARTIR DU GÉNOME

Une des principales utilisations du clonage est d'isoler les gènes spécifiques directement du génome. Un gène particulier ne représente qu'une infime partie du génome eucaryote. Par exemple, dans un mammifère typique, la taille du génome est d'environ 10^9 pb, ce qui signifie qu'un gène de 5 000 pb (par exemple) ne représente que 0,0005 % de l'ADN nucléaire total.

Afin d'identifier une si petite partie, il faut posséder une **sonde** très spécifique qui *ne réagit* qu'avec la séquence à laquelle on s'intéresse et qui sera capable de la reconnaître parmi une multitude d'autres séquences. Généralement, on utilise une sonde d'ADN ou d'ARN marquée radioactivement, et dont l'hybridation avec le gène est testée par autoradiographie.

Le point épineux est d'obtenir l'ARNm qui correspond à une protéine particulière, ce qui peut s'avérer être très difficile si le produit est rare. Il existe différentes techniques qui permettent d'isoler un ARNm grâce aux propriétés de son produit. L'une d'entre elles est la détection hautement sensible des protéines nouvellement synthétisées après injection de l'ARN dans l'oocyte de *Xenope*. Le problème avec ces techniques est qu'elles nécessitent que l'ARN soit déjà purifié.

La technique d'**arrêt de traduction par hybridation**, décrite sur la Figure 16.6, est une technique très puissante qui permet d'isoler directement l'ADNc complémentaire d'un ARNm. Elle permet l'identification d'un ADNc qui empêche la traduction d'un ARNm en s'hybridant à lui, ce qui a pour résultat la disparition du produit de cet ARN dans un système de traduction *in vitro***.

Un ADNc copie de l'ARNm est la meilleure des sondes pour isoler le gène correspondant. De larges quantités de matériel purifié peuvent être obtenues par clonage, matériel qui peut ensuite être marqué *in vitro* afin d'obtenir une radioactivité hautement spécifique.

La première étape pour identifier le gène correspondant à une sonde particulière est de couper l'ADN génomique en fragments de taille maniable. Il est désirable d'obtenir un gène en un minimum de fragments (idéalement, seulement un). Généralement, les longueurs maximales du génome qui peuvent être manipulables sont de l'ordre de 15-20 kb *. Il est parfois impossible d'obtenir un gène sous la forme d'un fragment unique, et sa structure doit ensuite être déterminée en faisant la synthèse des informations apportées par chacun des fragments.

La meilleure technique pour fragmenter le génome est de procéder à une digestion par une enzyme de restriction. Chaque fragment se termine par un site qui a été reconnu par une enzyme particulière. Toutefois, les sites de restriction ne sont pas forcément localisés à des endroits stratégiques ; ils peuvent se situer, par exemple, au milieu du gène à cloner. Pour éviter cela, on peut utiliser différentes enzymes de restriction, ce qui signifie qu'il faut répéter l'expérience avec plusieurs enzymes dont le site de reconnaissance se trouve à des endroits différents. Mais cela prend du temps, et lorsqu'il s'agit d'une longue séquence, il est difficile de trouver une enzyme qui ne la coupe pas.

Quand l'ADN d'un génome entier est digéré avec une enzyme de restriction, la fréquence de coupure est déterminée par la longueur de la séquence reconnue par l'enzyme. Plus la séquence est longue, moins il est probable qu'elle apparaisse fréquemment. Par exemple, la probabilité qu'une séquence particulière de 4pb apparaisse est de $0,25^4 = 1/256$, ce qui signifie qu'une enzyme à courte séquence de reconnaissance coupera l'ADN assez fréquemment. La fréquence passe à $1/1000$ pour une séquence de reconnaissance de 5 pb et à $1/4000$ pour une séquence de 6 pb.

(Ce calcul suppose que chaque base soit également représentée dans l'ADN, ce qui n'est généralement pas le cas. La fréquence de coupure peut être diminuée en utilisant une enzyme dont la séquence cible est constituée de paires de bases qui sont en moindre quantité dans la composition de l'ADN et *vice-versa*.)

Afin d'obtenir un ensemble de fragments de taille convenable à partir d'une digestion, on a recours à une ruse qui permet de réduire la fréquence de coupure. Une enzyme à courte séquence de reconnaissance (4 pb) est

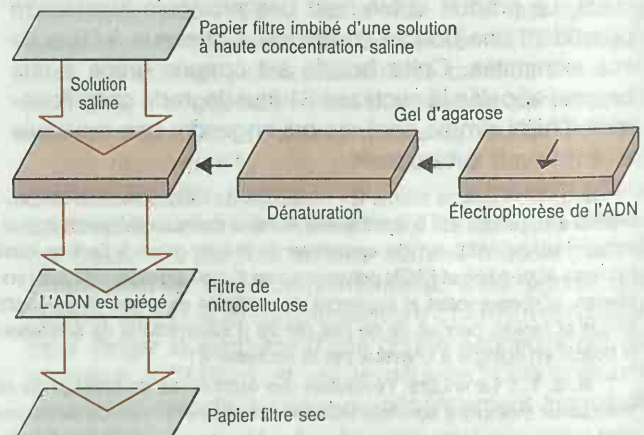
utilisée dans des conditions ne permettant qu'une digestion partielle de l'ADN. Les sites cibles ne sont coupés qu'occasionnellement et, par conséquent, la présence du site cible dans la séquence que l'on recherche ne nous empêche pas de l'obtenir intacte. Le fait que les sites apparaissent fréquemment dans le génome et que leur coupure soit rare, signifie que la distribution des fragments est proche d'une distribution obtenue par coupure au hasard dans le génome. Cependant, ici, les fragments se terminent tous par la même séquence, qui peut être une extrémité à bouts cohésifs, et qui sera facilement utilisable pour le clonage.

La distribution des sites reconnus par l'enzyme devient une question de hasard si l'on considère le génome dans son entier. La digestion de l'ADN eucaryote par une enzyme de restriction engendre une panoplie de fragments. Quand ils migrent sur gel par électrophorèse, ces fragments apparaissent sous la forme d'une traînée où n'est visible aucune bande distincte (à l'exception des séquences répétées, voir Chapitre 21). En revanche, si l'on possède la sonde spécifique, il est possible de détecter les séquences correspondant à cette sonde parmi cet ensemble de fragments.

Les fragments d'ADN ne peuvent être analysés directement à partir du gel d'agarose. Il faut donc, pour identifier des fragments, avoir la possibilité de transférer l'ADN sur un milieu où les réactions d'hybridation sont réalisables. L'ADN est dénaturé afin d'obtenir des fragments monocaténares qui seront ensuite transférés sur un filtre (de nitrocellulose ou de nylon le plus souvent)

Figure 16.7

Le « Southern blotting » permet le transfert de fragments d'ADN du gel d'électrophorèse sur un filtre de nitrocellulose.



* N. d. T. : Les vecteurs cosmiques peuvent accepter jusqu'à 35 kb. Une nouvelle méthode de clonage dans la levure décrite en 1987 permettrait de cloner des fragments de plusieurs centaines de kb.

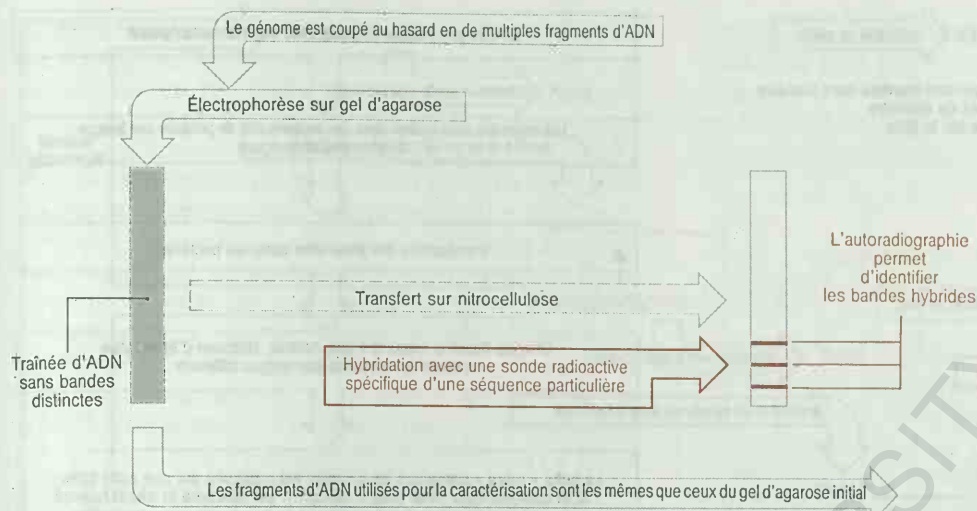


Figure 16.8
Le « Southern blotting » permet d'isoler les fragments d'ADN correspondant à une sonde particulière, directement à partir d'une digestion de l'ADN du génome.

où ils sont immobilisés. La technique utilisée pour transférer l'ADN est apparentée au « blotting » (buvardage), ce nom est utilisé couramment pour la décrire. Quand elle permet le transfert d'ADN, la technique s'appelle le « **Southern blotting** » (du nom de l'inventeur).

La **Figure 16.7** illustre le protocole de transfert de fragments d'ADN. Le gel d'agarose est placé sur un papier filtre qui a été préalablement imbibé d'une solution à haute concentration saline. Puis, on positionne le filtre de nitrocellulose sur le gel ; des papiers filtres secs sont ensuite mis en contact avec la nitrocellulose. La solution saline est attirée par le papier filtre sec. Pour arriver jusqu'à lui, la solution doit d'abord passer à travers le gel d'agarose puis à travers le filtre de nitrocellulose. Elle entraîne l'ADN à son passage, ADN qui sera ensuite emprisonné dans la nitrocellulose.

Sur la **Figure 16.8**, il est décrit qu'après avoir été immobilisé sur la nitrocellulose, l'ADN peut être hybridé *in situ* avec une sonde radioactive. Seuls les fragments qui sont complémentaires de la sonde utilisée s'hybrident avec elle. La sonde étant radioactive, l'hybridation pourra être visualisée par autoradiographie. Chaque séquence complémentaire de la sonde donnera lieu à une bande marquée radioactivement dont la position sur le filtre détermine la taille.

Cette technique est aussi utilisable avec l'ARN. Afin de transférer l'ADN de l'agarose vers un milieu adapté à une hybridation, il est nécessaire d'apporter quelques modifications à la technique. La technique modifiée pour les ARN est appelée le « **Northern blotting** » (quand elle est adaptée aux protéines, la technique prend le nom de « **Western blotting** »).

L'on désire parfois détecter la présence de séquences

particulières dans l'ARN d'une variété de types cellulaires. Adaptée à ce besoin, une variante des techniques de transfert d'ARN a été mise au point qui est appelée le « **dot blotting** ». Comme l'illustre la **Figure 16.9**, des ADN contenant les séquences étudiées (ADNc ou génomiques) sont déposés sous forme de cercles (« dot ») adjacents sur le filtre. Le filtre est ensuite hybridé avec les ARN marqués radioactivement des cellules analysées. Si les cellules contiennent des ARN correspondant aux séquences déposées sur le filtre, celles-ci apparaîtront sous la forme de tâches radioactives. L'intensité de ces tâches dépend du niveau d'expression des ADN correspondants dans les cellules étudiées.

LA MARCHÉ SUR LE CHROMOSOME

L'isolement de fragments, directement à partir d'une digestion génomique, pose des problèmes pratiques. Un moyen puissant pour obtenir un gène particulier est d'inverser l'ordre des événements, c'est-à-dire de cloner le génome en premier puis de sélectionner les clones contenant la séquence recherchée. Les vecteurs véhiculant de l'ADN issu du génome sont appelés des **clones d'ADN génomiques** ou **chromosomiques** (en opposition aux clones d'ADNc qui représentent des ARNm).

Le clonage d'un génome entier (à opposer au clonage de fragments spécifiques) est souvent appelé l'**expérience de « shotgun »** (coup de fusil). En premier lieu, le génome est cassé en fragments de taille maniable. Puis les fragments sont insérés dans un vecteur de clonage afin d'engendrer une population de vecteurs chimériques. Un ensemble de fragments clonés de la

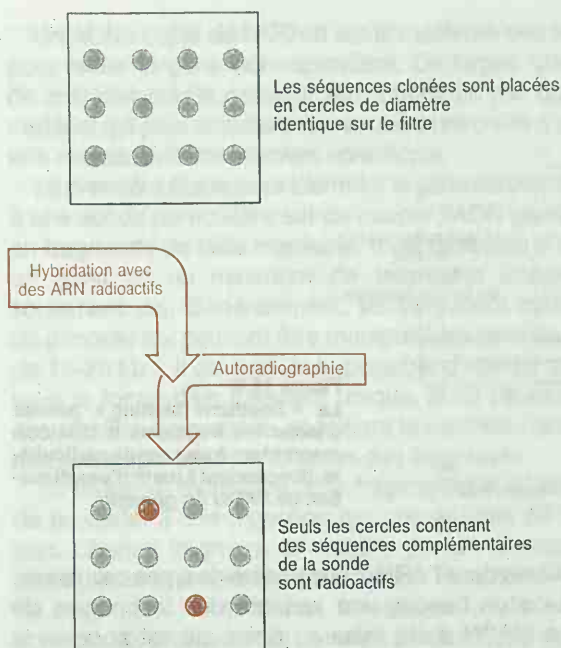


Figure 16.9

Le « dot blotting » permet de tester de nombreuses séquences clonées quant à leur homologie avec une sonde marquée radioactivement. Les séquences clonées sont déposées sous forme de cercles sur un papier filtre qui est hybridé avec une solution contenant la sonde radioactive. Les cercles contenant les séquences complémentaires à la sonde apparaissent par autoradiographie.

sorte est appelé une **banque génomique**. Une fois la banque réalisée au moyen d'un vecteur phagique ou plasmidique (plus souvent phagique car il est plus facile de cloner le nombre nécessaire de molécules d'ADN chimériques grâce à un phage), elle peut être perpétuée indéfiniment et facilement réutilisée lorsqu'une nouvelle sonde est disponible pour rechercher un fragment particulier.

Le nombre de fragments à cloner pour que *chaque* séquence du génome soit représentée dans au moins un plasmide chimérique, varie en fonction de la taille des fragments, de la taille du génome et de la probabilité désirée. Pour obtenir un taux de probabilité de 99 % de représentativité de fragments de 15 à 20 kb, il faut cloner 1 500 fragments chez *E. coli*, 4 600 chez la levure, 48 000 chez *D. melanogaster* et 800 000 chez les mammifères. Dans chacun de ces cas, une telle banque a été réalisée.

Comment peut-on sélectionner un clone génomique particulier de la banque ? La technique d'**hybridation sur colonies** est illustrée sur la Figure 16.10. Les colonies bactériennes portant des vecteurs chimériques sont lysées sur des filtres de nitrocellulose. Leur ADN est

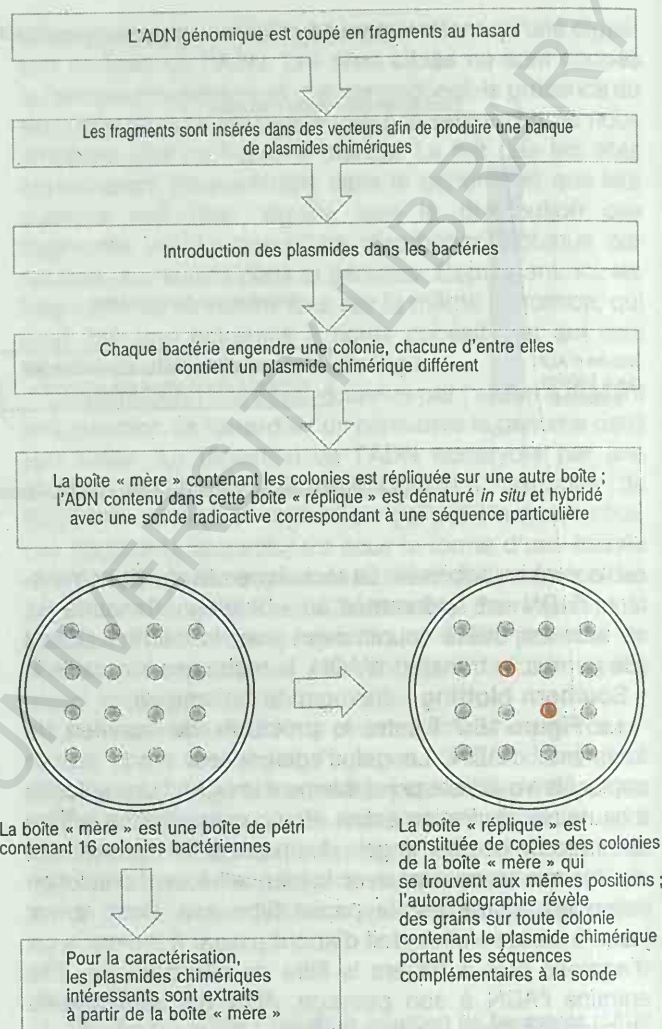


Figure 16.10

L'hybridation sur colonies permet de sélectionner un plasmide chimérique portant une séquence particulière grâce à sa complémentarité avec une sonde marquée radioactivement.

ensuite dénaturé *in situ* et fixé sur le filtre. Le filtre est hybridé avec une sonde radioactive qui correspond à la séquence recherchée (généralement un clone d'ADNc cloné). Toute colonie où se produit l'hybridation est visualisée sous forme d'une tache noire par autoradiographie. Les vecteurs chimériques correspondants peuvent ensuite être isolés grâce à la banque de référence utilisée.

Un clone génomique peut être porteur seulement d'une partie de la séquence de la sonde pour réagir avec elle (le minimum requis est habituellement d'environ 50 pb). En effet, quand un gène eucaryote est long, il a des chances d'être fragmenté au moment de la fabrication

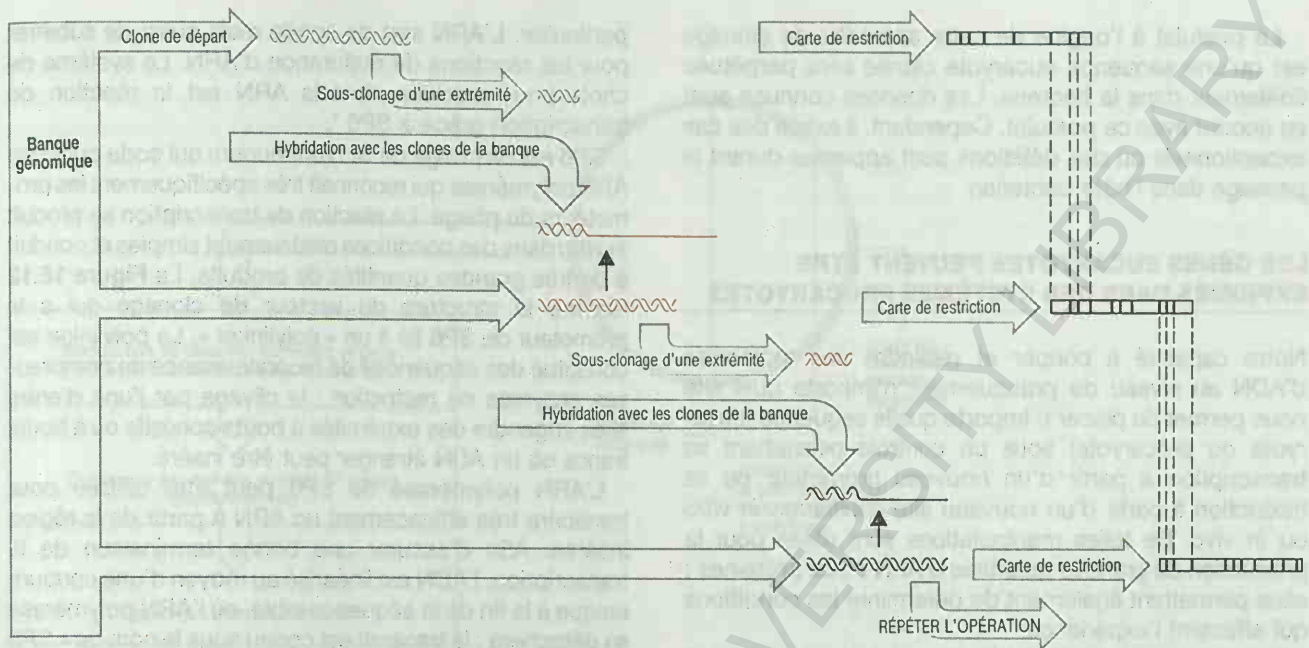


Figure 16.11
Des hybridations successives entre clones génomiques chevauchants permettent de marcher sur le chromosome.

de la banque. Dans le cas d'une digestion partielle, la coupure au hasard engendre des fragments qui se chevauchent, car, de part et d'autre d'une séquence donnée dans plusieurs génomes, différents sites ont été attaqués par la digestion. Si aucun des clones génomiques ne contient la séquence recherchée dans son entier, il est nécessaire de reconstruire la séquence génomique initiale, ce qui est rendu possible en utilisant les chevauchements qui existent entre chaque fragment.

Le principe d'identification de fragments chevauchants permet la reconstruction de longs gènes, mais aussi la caractérisation de longues régions du chromosome. Un des points intéressants concernant le génome eucaryote est de connaître la nature du contexte dans lequel réside un gène ainsi que ses relations avec les gènes avoisinants. Afin de caractériser une région par-delà le gène lui-même, l'on s'aventure sur des « terres » inconnues dans le sens qu'aucune sonde dérivée de produits de gène ne les reconnaît. Chaque fragment du génome est isolé uniquement en vertu du fait qu'il possède un fragment qui chevauche un autre fragment.

Cette technique est appelée la **marche sur le chromosome** et est illustrée sur la **Figure 16.11**. Le point de départ est un clone qui a pu être isolé parce qu'il contient un gène connu ou parce que nous savons (par cartographie génétique) qu'il est proche de la région qui nous intéresse. Ce fragment pouvant être assez important, il est plus facile d'en utiliser un sous-fragment correspondant à l'une de ses extrémités, pour isoler le suivant. Ensuite, sont identifiés les autres vecteurs chimériques de la banque, chevauchant avec ce clone. Les nouveaux vecteurs étendront le fragment porté par le premier clone isolé, dans un sens ou dans l'autre. On peut déterminer la direction de cette extension en effectuant la carte de restriction de chaque fragment. Ce procédé peut être répété indéfiniment.

En principe, il est possible de marcher sur des centaines de kb ; des régions du génome voisines de 1 000 kb ont été cartographiées de cette façon. Il est généralement possible de marcher de l'ordre de 100 kb par mois, ce qui est pratique si l'on désire caractériser de longues régions*.

* **N. d. T. :** Les performances annoncées par l'auteur ont maintenant été largement dépassées. Des régions de plusieurs milliers de kb ont été précisément cartographiées, par exemple dans la région de l'immense gène de la dystrophine, gène dont l'altération entraîne la myopathie de Duchenne de Boulogne. L'utilisation des vecteurs cosmiques, et peut-être demain de fragments plus grands encore, clonés dans la levure, permettent de chausser des bottes de sept lieux pour cette marche sur le chromosome. Cette cartographie de larges régions du génome est également facilitée grâce à l'utilisation d'une nouvelle méthode de séparation de longs segments d'ADN par électrophorèse en champ pulsé. Cette méthode consiste à faire migrer des fragments d'ADN allant jusqu'à 2 000 à 3 000 kb dans un champ électrique dont le sens est périodiquement inversé.

Le postulat à l'origine de cette approche de clonage est qu'une séquence eucaryote clonée sera perpétuée fidèlement dans la bactérie. Les données connues sont en accord avec ce postulat. Cependant, il existe des cas exceptionnels où des délétions sont apparues durant le passage dans l'hôte bactérien.

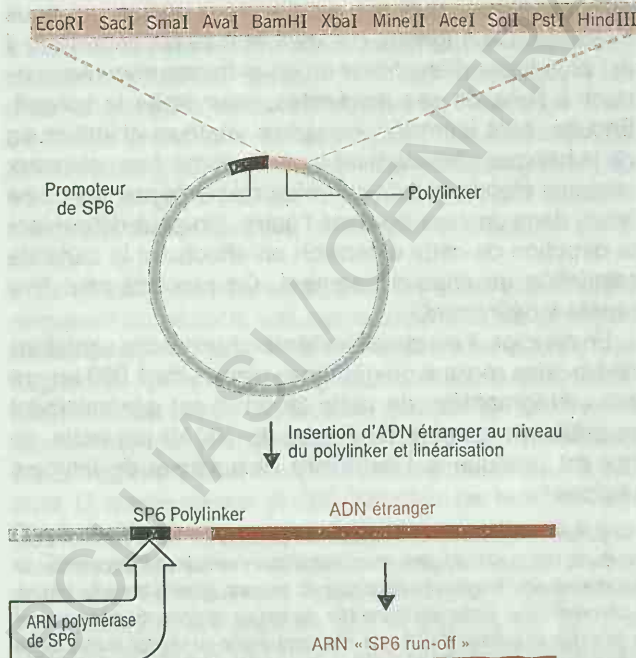
LES GÈNES EUCARYOTES PEUVENT ÊTRE EXPRIMÉS DANS DES SYSTÈMES PROCARYOTES

Notre capacité à couper et rejoindre des molécules d'ADN au niveau de pratiquement n'importe quel site nous permet de placer n'importe quelle séquence (eucaryote ou procaryote) sous un contrôle permettant sa transcription à partir d'un nouveau promoteur ou sa traduction à partir d'un nouveau site d'initiation *in vitro* ou *in vivo*. De telles manipulations sont utiles pour la production de grandes quantités d'ARN ou de protéines ; elles permettent également de déterminer les conditions qui affectent l'expérience.

Il est souvent souhaitable de produire de grandes quantités d'ARN simple brin correspondant à un gène

Figure 16.12

Toute séquence d'ADN insérée dans la région polylinker peut être transcrite en un ARN simple brin à partir du promoteur de SP6 et grâce à l'ARN polymérase de SP6.



* N. d. T. : Sont également utilisées les ARN polymérases des phages T3 et T7.

particulier. L'ARN sert de sonde mais aussi de substrat pour les réactions de maturation d'ARN. Le système de choix pour produire de tels ARN est la réaction de transcription grâce à SP6*.

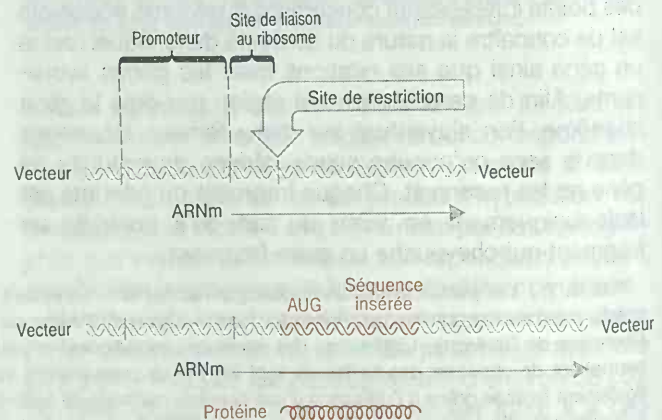
SP6 est un phage de *S. typhimurium* qui code pour une ARN polymérase qui reconnaît très spécifiquement les promoteurs du phage. La réaction de transcription se produit *in vitro* dans des conditions relativement simples et conduit à de très grandes quantités de produits. La Figure 16.12 résume la structure du vecteur de clonage qui a le promoteur de SP6 lié à un « polylinker ». Le polylinker est constitué des séquences de reconnaissance de nombreuses enzymes de restriction ; le clivage par l'une d'entre elles engendre des extrémités à bouts cohésifs ou à bouts francs où un ADN étranger peut être inséré.

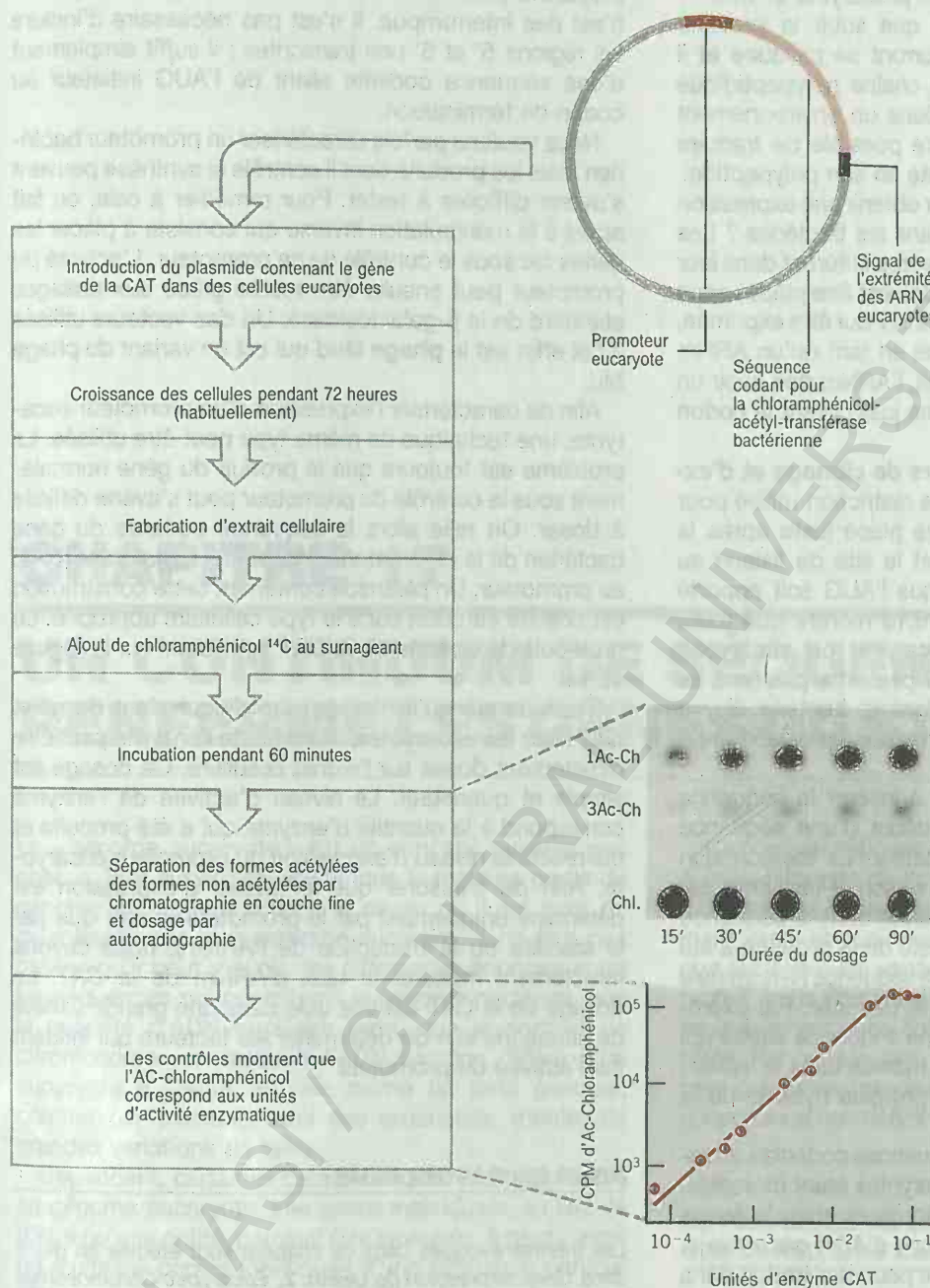
L'ARN polymérase de SP6 peut être utilisée pour transcrire très efficacement un ARN à partir de la région insérée. Afin d'assurer une bonne terminaison de la transcription, l'ADN est linéarisé au moyen d'une coupure unique à la fin de la séquence cible, où l'ARN polymérase se détachera ; le transcrit est connu sous le nom de « SP6 run-off ». L'ARN peut être produit en grande quantité et hautement marqué si l'on utilise un nucléotide précurseur radioactif. L'ARN peut également être traduit *in vitro* si l'on désire obtenir la protéine.

Le code génétique étant universel, une séquence codante donnée a toujours la même signification. (Il existe des exceptions dans les génomes mitochondriaux ; voir Chapitre 6.) Par conséquent, si une séquence intacte codant pour une protéine eucaryote est portée sur un vecteur chimérique, il devrait être possible de la transcrire

Figure 16.13

Toute séquence codante eucaryote peut être transcrite si elle est insérée dans un site approprié d'un vecteur de clonage.



**Figure 16.14**

Le dosage de la CAT peut être utilisé pour suivre l'activité de tout promoteur eucaryote. L'activité de la séquence codante portée par le plasmide est dosée par mesure de la capacité de l'extrait cellulaire à convertir le chloramphénicol marqué radioactivement en sa forme acétylée. La proportion de chloramphénicol acétylé augmente en fonction du temps et est directement proportionnelle au nombre d'unités d'enzyme présente.

en un ARNm qui pourra lui-même être traduit dans un hôte bactérien ou par un système procaryote *in vitro*.

Cependant, les modifications que subit la protéine dans son habitat naturel ne pourront se produire et il reste, bien sûr, le risque que la chaîne polypeptidique soit instable dans la bactérie. Dans un environnement approprié, il devrait toutefois être possible de traduire toute séquence codante eucaryote en son polypeptide.

Quelles sont les conditions pour obtenir une expression correcte de gènes eucaryotes dans les bactéries ? Les promoteurs eucaryotes et procaryotes différant dans leur séquence, les gènes eucaryotes doivent être placés sous le contrôle de promoteurs bactériens. Pour être exprimée, l'insertion eucaryote est transcrite en tant qu'un ARNm monocistronique. Pour être traduit, l'ARNm doit avoir un site correct de liaison au ribosome juste avant le codon AUG initiateur.

Ont été construits des **vecteurs de clonage et d'expression** dans lesquels le site de restriction utilisé pour insérer l'ADN étranger, se trouve placé juste après la région contenant le promoteur et le site de liaison au ribosome. L'on s'arrange pour que l'AUG soit apporté par l'ADN étranger. La **Figure 16.13** montre que toute séquence d'ADN étranger (procaryote ou eucaryote) commençant par un codon AUG peut être placée à ce site puis transcrite et traduite dans la bactérie. Il y a production d'une protéine qui correspond précisément à la région codante insérée.

Une autre technique consiste à insérer la séquence codante tout de suite après le début d'une séquence codant pour une protéine du vecteur. La transcription produit une protéine hybride ; la région N-terminale est constituée de la protéine bactérienne dont la séquence a été interrompue alors que le reste de la protéine a été codé par la séquence étrangère. La séquence N-terminale bactérienne peut être très courte et très utile. Par exemple, elle peut être constituée d'une séquence signal qui assure la sécrétion de la protéine hybride dans le milieu ; elle peut également protéger la protéine hybride de la dégradation dans la bactérie.

D'où doivent être issues les séquences codantes eucaryotes ? De nombreux gènes eucaryotes étant morcelés, la séquence codante n'est pas toujours sous la forme d'une région continue dans un clone d'ADN génomique. Dans ce cas, l'ADN génomique ne peut être traduit dans la bactérie. Par conséquent, il est préférable d'utiliser une

séquence codante isolée à partir d'un clone d'ADNc préparé à partir de l'ARNm où là, la séquence codante n'est pas interrompue. Il n'est pas nécessaire d'inclure les régions 5' et 3' non transcrites ; il suffit simplement d'une séquence codante allant de l'AUG initiateur au codon de terminaison.

Nous voulons parfois caractériser un promoteur bactérien mais les produits dont il contrôle la synthèse peuvent s'avérer difficiles à tester. Pour remédier à cela, on fait appel à la manipulation inverse qui consiste à placer les gènes *lac* sous le contrôle de ce promoteur. L'activité du promoteur peut ensuite être testée grâce aux dosages standard de la β -galactosidase. Un des vecteurs utilisés à cet effet est le **phage Mud** qui est un variant du phage Mu.

Afin de caractériser l'expression d'un promoteur eucaryote, une technique de même type peut être utilisée. Le problème est toujours que le produit du gène normalement sous le contrôle du promoteur peut s'avérer difficile à doser. On relie alors la séquence codante du gène bactérien de la *chloramphénicol acétyl transférase* (CAT) au promoteur. Un plasmide contenant cette construction est ensuite introduit dans le type cellulaire approprié. Le protocole du **dosage de la CAT** est illustré sur la **Figure 16.14**.

Étant donné qu'il n'existe pas d'équivalent du gène CAT chez les eucaryotes, l'activité de l'enzyme peut être directement dosée sur l'extrait cellulaire. Ce dosage est simple et quantitatif. Le niveau d'activité de l'enzyme correspond à la quantité d'enzyme qui a été produite et qui reflète le niveau d'expression du promoteur eucaryote. Afin de s'assurer que le niveau d'expression est déterminé uniquement par le promoteur (plutôt que par la stabilité ou la traduction de l'ARNm), nous devons directement mesurer le taux d'ARNm de la CAT. Le dosage de la CAT est très utile dans une grande variété de situations afin de déterminer les facteurs qui influent sur l'activité du promoteur*.

POUR EN SAVOIR PLUS

Les thèmes évoqués dans ce chapitre sont étudiés en détail dans *Gene expression* de Lewin, 2, *Eucaryotic Chromosomes* (Wiley, New York, 1980, pp. 761-765).

* N. d. T. : Le gène CAT est appelé habituellement gène rapporteur. D'autres gènes rapporteurs sont également très utilisés, tel celui de la luciférase de ver luisant ou des gènes de résistance à des antibiotiques actifs sur des cellules eucaryotes.

CHAPITRE 17

UN CONTINUUM DE SÉQUENCES RENFERME LES GÈNES DE STRUCTURE

Le « génome eucaryote » est un concept quelque peu obscur. Être eucaryote signifie que la majeure partie du génome est contenue dans le noyau où il est isolé du cytoplasme par la membrane nucléaire. Il ne faut pas oublier qu'il peut y avoir des différences énormes entre les séquences qui constituent le génome nucléaire, que la quantité d'ADN nucléaire varie, que le nombre de chromosomes en lesquels l'ADN ségrège diffère d'un eucaryote à l'autre, et que même un petit génome, comme par exemple celui des organelles, montre de grandes variations de taille.

Cependant, certaines caractéristiques restent propres au génome eucaryote. Les gènes individuels, au lieu de n'être qu'une entité, peuvent être morcelés. Il peut y avoir de multiples copies identiques d'une séquence particulière et de grandes régions d'ADN qui ne codent pour aucune protéine. A cause de cette barrière qui sépare noyau et cytoplasme, l'expression d'un gène chez les eucaryotes met en œuvre des dispositifs différents de ceux utilisés par les procaryotes.

Ces caractéristiques nous permettent-elles de conclure à une uniformité de l'organisation de l'ADN eucaryote ? Il ne faut pas oublier que le règne eucaryote est extrêmement vaste et que nous ne possédons des informations que sur quelques espèces. En étudiant l'ADN des eucaryotes, nous décrivons des caractéristiques qui existent probablement dans chaque génome à des degrés différents. Par conséquent, nous devons établir les mécanismes possibles responsables de l'organisation de l'ADN eucaryote et non définir un modèle hypothétique général.

LE PARADOXE DE LA VALEUR C DÉCRIT DES VARIATIONS DANS LA TAILLE DU GÉNOME

La quantité totale d'ADN dans le génome (haploïde) est une caractéristique de chaque espèce vivante, connue comme étant sa **valeur C**. Il existe d'énormes variations de cette valeur qui peut aller d'aussi peu que 10^6 pb dans

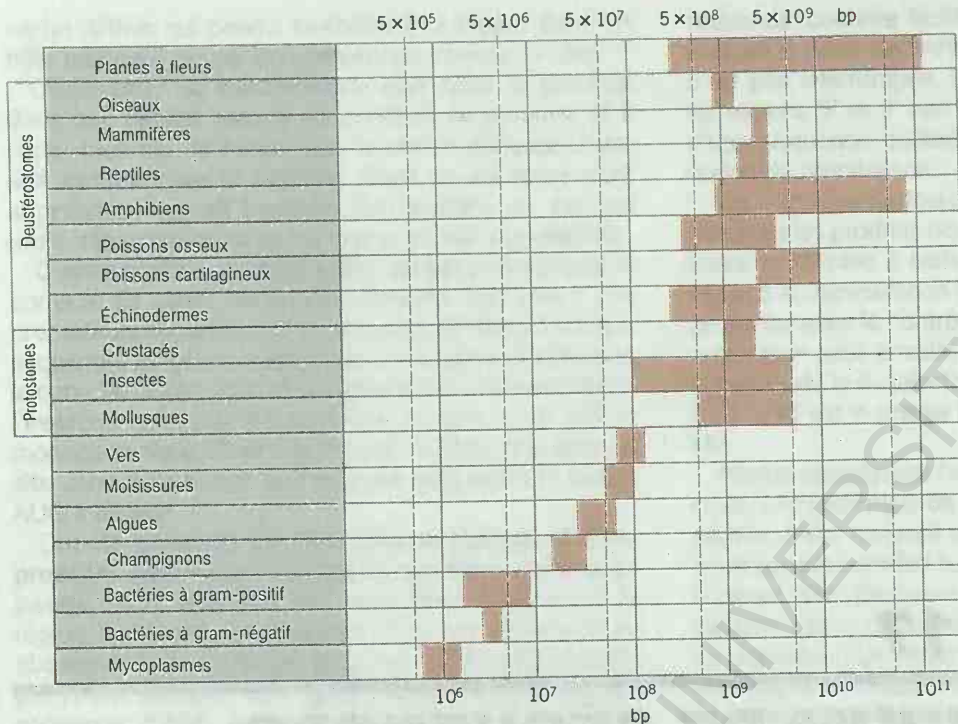


Figure 17.1

La quantité d'ADN d'un génome haploïde n'est pas le reflet direct de la complexité morphologique des espèces. La taille du génome augmente des procaryotes aux protostomes, mais varie largement chez les deutérostomes. La variation des valeurs d'ADN à l'intérieur d'un phylum est représentée par une zone sombre.

le cas d'un mycoplasme à 10¹¹ pb chez certaines plantes et amphibiens.

Les valeurs C trouvées dans différents phyla de l'évolution sont résumées sur la Figure 17.1. La taille minimum d'un génome augmente dans chaque phylum avec la complexité du génome.

Les eucaryotes unicellulaires (dont le mode de vie ressemble à celui des procaryotes) ont de petits génomes bien que plus importants que ceux des bactéries. Être eucaryote ne signifie pas que la taille du génome est largement augmentée ; par exemple, la taille du génome de la levure *S. cerevisiae* est d'environ 2,3 × 10⁷ pb, seulement cinq fois plus grande que celle de la bactérie *E. coli*.

Une augmentation de deux fois de la taille du génome est suffisante pour la moisissure de vase *D. discoideum* qui est capable de vivre aussi bien sous le mode unicellulaire que multicellulaire. La taille du génome doit encore

augmenter pour qu'apparaissent les premiers organismes totalement multicellulaires. Par exemple, le ver nématode *C. elegans* a un génome de 8 × 10⁷ pb.

Si l'on remonte plus avant l'arbre de l'évolution, la relation qui unit la complexité d'un organisme et la taille de son génome devient obscure, bien qu'il soit nécessaire d'avoir une taille de génome de plus de 10⁸ pb pour produire un insecte, plus de 4 × 10⁸ pb pour un échinoderme, plus de 8 × 10⁸ pb pour un oiseau ou un amphibien, et plus de 2 × 10⁹ pb pour un mammifère.

Dans certains cas, l'échelle de variation de la taille du génome est faible. Par exemple, les oiseaux, les reptiles et les mammifères montrent tous peu de variations, au sein d'un phylum, dans la taille du génome qui passe au maximum du simple au double dans chacun des cas. Mais dans d'autres cas, il existe une grande variation dans la valeur C qui peut augmenter jusqu'à dix fois dans le même phylum.

Le **paradoxe de la valeur C** tire son nom de notre incapacité à définir le génome en terme de fonctions connues.

Une caractéristique étrange est l'existence de grandes variations entre les valeurs C d'espèces dont l'apparente complexité semble identique. Une telle variation est trouvée chez les amphibiens où les plus petits génomes sont juste inférieurs à 10^9 pb alors que les plus grands sont presque de 10^{11} pb. Il est difficile de croire que cette variation reflète une différence de 100 fois dans le nombre de gènes nécessaires à la création de différents amphibiens.

De surcroît, des espèces très proches montrent des différences surprenantes dans la taille de leur génome. Par exemple, deux espèces d'amphibiens dont les morphologies sont très similaires peuvent avoir une quantité relative d'ADN différant d'un facteur de 10 l'un par rapport à l'autre. Il semble improbable, encore une fois, qu'il y ait une différence de cet ordre dans le nombre de gènes. Si le nombre de gènes est approximativement identique, une grande partie de l'ADN des espèces possédant un génome important ne code pas pour des protéines : quelle peut donc être sa fonction ?

Un autre aspect du paradoxe de la valeur C est l'excès apparent d'ADN par rapport à la quantité attendue si l'on s'intéresse aux protéines. Nous savons maintenant qu'une partie de cet ADN « en excès » rend compte du fait que les gènes sont bien plus importants que les séquences nécessaires pour coder les protéines (principalement, à cause des séquences qui s'interposent dans la région codante et la morcelent). Cette raison n'est toutefois pas suffisante pour rendre compte de cette importante quantité d'ADN.

Même si l'on suppose que la taille moyenne d'un gène est de 5 000 à 8 000 pb (ce qui est plus important que la plupart des gènes morcelés connus), le nombre de gènes dans le génome des mammifères serait de 400 000 à 600 000. Ce chiffre est-il plausible ?

Il ne l'est probablement pas : bien que le nombre de gènes ne soit pas connu avec précision, il semble qu'il faille les compter en dizaines de mille plutôt qu'en centaines de mille.

Une estimation du nombre de gènes peut être effectuée en tenant compte du nombre d'ARNm différents (voir plus loin). Bien que cela ne soit pas réalisable pour chaque type cellulaire (pour ensuite calculer une somme totale pour l'organisme), il semble que le nombre de gènes exprimés dans un type cellulaire donné soit d'environ 10 000. La plupart de ces gènes sont communs à la plupart ou toutes les cellules de l'organisme. Par consé-

quent, cette valeur est probablement à multiplier par 2-4 pour avoir le nombre total de gènes exprimés. En laissant une incertitude sur l'estimation du nombre de gènes présents à l'état de multiples copies, nous pouvons dire que le génome des mammifères semble avoir environ 30 000-40 000 gènes fonctionnels.

Un moyen indirect pour tenter d'estimer le nombre de gènes essentiels est d'identifier tous les loci pouvant être mutés. Ce travail a été effectué chez *D. melanogaster* et a permis d'évaluer à environ 5 000 le nombre de gènes essentiels chez cet insecte. En estimant raisonnablement à 2 000 pb la taille d'un gène, cela correspond à un génome de 10^7 pb, 10 fois moins important que la quantité d'ADN présente.

Bien sûr, les gènes identifiés par mutation sont ceux pour lesquels une altération a des effets visibles ou léthaux. Il est possible que seuls certains gènes entrent dans cette catégorie. Cela impliquerait qu'une grande proportion des gènes, voire la majorité, codent pour des protéines non essentielles à la survie de l'organisme (dans le sens qu'une altération mutationnelle de ces gènes ne produit pas d'effet détectable).

Il reste plusieurs questions clés à résoudre. Quelle est la proportion d'ADN du génome qui représente réellement les protéines, c'est-à-dire quelle est la proportion d'ADN correspondant aux gènes codant directement pour les protéines, aux régions s'interposant entre les parties codantes et aux régions flanquantes transcrites ? Sur le nombre total de gènes, combien sont essentiels et combien sont accessoires ? Quelle est la fonction (s'il y en a une) de l'ADN où ne réside aucun gène ? Quel effet peut avoir un important changement de taille sur le fonctionnement du génome, comme c'est le cas chez certains amphibiens ?

LES CINÉTIQUES DE RÉASSOCIATION DÉPENDENT DE LA COMPLEXITÉ DE LA SÉQUENCE

La réassociation entre des séquences complémentaires d'ADN s'effectue par appariement de bases, en opposition à la dénaturation qui les a séparées (voir Figure 3.2). Nous avons déjà mentionné un grand nombre d'expériences qui reposent, par extension, sur la propriété de réassociation de l'ADN puisqu'elles permettent d'isoler des séquences d'ADN ou d'ARN grâce à leur capacité à s'hybrider avec une sonde particulière. Nous allons voir maintenant que les différences de cinétique dans la réaction de réassociation reflètent des différences qui existent entre les séquences présentes, et que la réaction

peut être utilisée pour quantifier des gènes et les ARN qui leur correspondent. Quand elles sont effectuées en solution, de telles réactions sont décrites comme étant des **hybridations en milieu liquide**.

La renaturation de l'ADN dépend de la rencontre des brins complémentaires qui s'effectue au hasard, et suit des cinétiques de second ordre. La vitesse de la réaction est gouvernée par l'équation suivante :

$$\frac{dC}{dt} = -kC^2 \quad (1)$$

où C représente la concentration d'ADN monocaténaire au temps t , et k la constante de vitesse de réassociation.

En prenant la concentration initiale d'ADN, C_0 , au temps $t = 0$, et la concentration C d'ADN restant monocaténaire après le temps t , la progression de la réaction peut être décrite comme suit :

$$\frac{C}{C_0} = \frac{1}{1 + k.C_0 t_{1/2}} \quad (2)$$

Par conséquent, quand la réaction s'est produite à moitié, au temps $t_{1/2}$:

$$\frac{C}{C_0} = \frac{1}{2} = \frac{1}{1 + k.C_0 t_{1/2}} \quad (3)$$

et donc :

$$C_0 t_{1/2} = \frac{1}{k} \quad (4)$$

La réassociation de toute molécule d'ADN est caractérisée par une constante de vitesse k (dont l'unité est en moles⁻¹ de nucléotides par litre-sec.⁻¹), ou par sa valeur de $C_0 t_{1/2}$ (exprimée en moles de nucléotides × sec/litre).

L'équation 4 montre que le paramètre contrôlant la réaction de réassociation est le produit de la concentration d'ADN (C_0) et du temps d'incubation (t), que l'on décrit couramment comme étant le **Cot**. Quand la réassociation s'est produite à moitié, cette valeur est appelée le **Cot_{1/2}**. Le Cot_{1/2} étant le produit de la concentration et du temps requis pour que la réaction se soit effectuée à moitié, cela signifie que plus la valeur de Cot_{1/2} est élevée, plus la réaction est lente.

La **courbe de Cot** permet de suivre la réassociation de l'ADN, elle décrit la fraction d'ADN qui s'est réassociée ($1 - C/C_0$) en fonction du log du Cot. Sur la **Figure 17.2** sont dessinées les courbes de Cot obtenues pour différents génomes. L'allure des courbes est identique, c'est-à-dire que, pour chaque génome, la renaturation s'effectue à des valeurs de Cot qui varient d'environ 100 fois entre le moment où la réaction s'est produite à 10 % et celui où elle s'est produite à 90 %. Toutefois, le Cot requis dans chacun des cas est différent. Il est donné par le Cot_{1/2}.

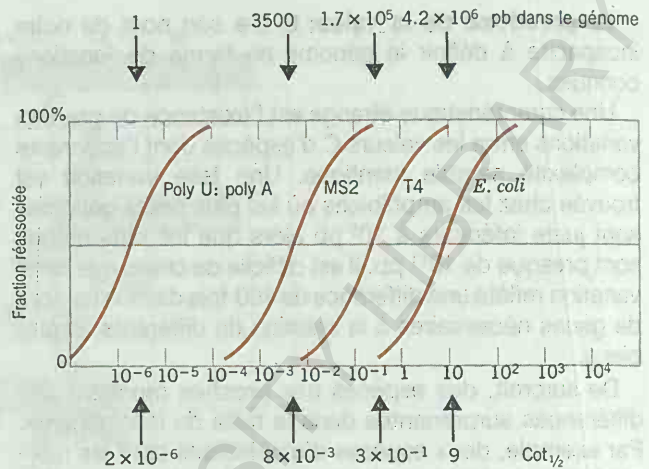


Figure 17.2

La vitesse de réassociation est inversement proportionnelle à la longueur de l'ADN se réassociant.

Quand elles sont déterminées par la technique de réassociation d'ADN, les quantités d'ADN sont généralement données en paires de bases ou en daltons. Quand elle est mesurée chimiquement, la quantité d'ADN est donnée en picogrammes (pg). L'équivalence est la suivante : 1 pg = 0,965 × 10⁹ pb = 6,1 × 10¹¹ daltons.

Le Cot_{1/2} est directement proportionnel à la quantité d'ADN dans le génome. Cela reflète le fait que, plus le génome devient complexe, moins il y a de copies d'une séquence particulière dans une masse donnée d'ADN. Par exemple, si la C_0 est de 12 pg, cela signifie que dans un génome bactérien dont la taille est de 0,004 pg, il y a 3 000 copies de chaque séquence, mais que, dans un génome eucaryote dont la taille est de 3 pg, il y a seulement 4 copies de chaque séquence. Par conséquent, pour une même concentration *absolue* d'ADN mesurée en moles de nucléotides par litre (c'est-à-dire la C_0) chez un eucaryote et chez une bactérie, la concentration de chaque séquence eucaryote est 750 fois plus basse que celle de chaque séquence bactérienne.

La vitesse de réassociation étant dépendante de la concentration en séquences complémentaires, il doit y avoir 750 fois plus d'ADN dans ce génome eucaryote si l'on veut que chaque séquence eucaryote soit présente à la même concentration relative que chaque séquence bactérienne (ou alors, il faut incuber la même quantité d'ADN 750 fois plus longtemps). Par conséquent, le Cot_{1/2} de la réaction du génome eucaryote est 750 fois le Cot_{1/2} de la réaction bactérienne.

En fait, le Cot_{1/2} de la réaction indique la *longueur totale des différentes séquences* qui sont présentes. Ceci est

décrit comme étant la **complexité**, qui est généralement donnée en paires de bases.

Par conséquent, la renaturation de l'ADN de tout génome (ou d'une partie du génome) est caractérisée par un $Cot_{1/2}$ qui est proportionnel à la complexité du génome. Ainsi, la complexité de tout ADN peut être déterminée en comparant son $Cot_{1/2}$ à celui d'un ADN standard de complexité donnée. On utilise généralement comme ADN standard l'ADN d'*E. coli* dont la complexité est considérée comme étant identique à la taille du génome (ce qui signifie que chaque séquence du génome d'*E. coli* qui est de $4,2 \times 10^6$ pb, est unique). Nous pouvons donc écrire l'équation suivante :

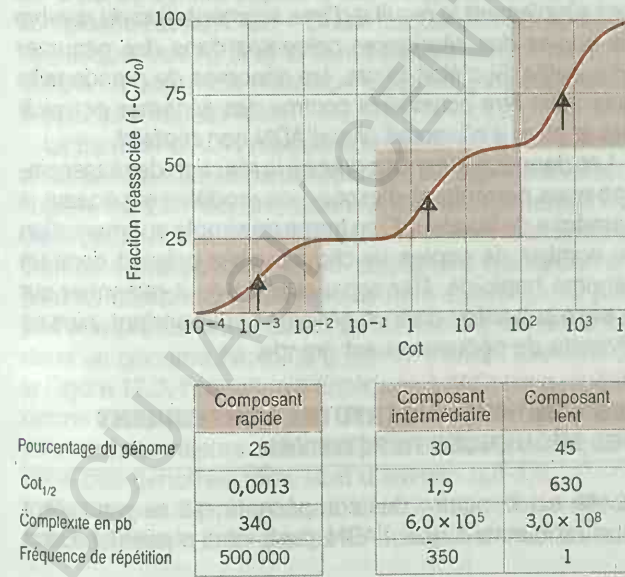
$$\frac{Cot_{1/2} \text{ (ADN de tout génome)}}{Cot_{1/2} \text{ (ADN d' } E. coli \text{)}} = \frac{\text{Complexité de tout génome}}{4,2 \times 10^6 \text{ pb}}$$

(5)

LES GÉNOMES EUCARYOTES CONTIENNENT, EN TERME DE SÉQUENCES, DES COMPOSANTS DIFFÉRENTS

Si l'on caractérise l'ADN d'un génome eucaryote par ses cinétiques de réassociation, la réaction se produit généralement à des valeurs de Cot qui varient d'un ordre de grandeur de 1 à 8. Nous sommes loin des variations de 100 fois attendues de l'équation 2 et décrites dans les exemples de la Figure 17.2. La raison en est que, dans l'équation, on a considéré qu'il n'y a qu'un seul composant de **cinétique pure** dans la réaction de réassociation.

Figure 17.3
Les cinétiques de réassociation de l'ADN eucaryote mettent en évidence trois types de composants (indiqués par les zones sombres).



En réalité, un génome est constitué de plusieurs composants, chacun possédant ses **propres caractéristiques cinétiques**.

La **Figure 17.3** décrit la réassociation d'un génome (hypothétique) eucaryote, commençant à un Cot de 10^{-4} et se terminant à un Cot de 10^4 . La réaction est découpée en trois phases distinctes, représentées par les zones sombres. Un plateau sépare les deux premières phases, tandis que la seconde et la troisième phases se chevauchent légèrement. Chacune de ces phases correspond à la réassociation d'un composant du génome de cinétique différente.

La première fraction qui se réassocie est appelée le **composant rapide**. Ici, elle représente 25 % de l'ADN total et se renature à des valeurs de Cot allant de 10^{-4} à environ 2×10^{-2} , avec une valeur de $Cot_{1/2}$ de 0,0013.

La fraction suivante est appelée le **composant intermédiaire**. Elle représente 30 % de l'ADN. Elle se renature à des valeurs de Cot allant d'environ 0,2 à 100, avec un $Cot_{1/2}$ de 1,9.

La dernière fraction qui se renature est appelée le **composant lent**. Elle représente 45 % de l'ADN total et s'étend d'une valeur de Cot allant d'environ 100 à environ 10 000, avec un $Cot_{1/2}$ de 630.

Afin de calculer la complexité de ces fractions, elles doivent être considérées indépendamment les unes des autres comme des composants de cinétique différente ; leur réassociation sera comparée séparément avec celle d'un ADN standard. Le composant lent représente 45 % de l'ADN total, sa concentration dans la réaction de réassociation est 0,45 de la C_0 mesurée (qui se rapporte à la quantité totale d'ADN mis dans la réaction). Par conséquent, le $Cot_{1/2}$ de la fraction lente est $0,45 \times 630 = 283$.

Ainsi, si l'ADN lent avait été isolé des autres fractions comme un composant pur, il se serait renaturé avec un Cot de 283. Supposons que dans de telles conditions, l'ADN d'*E. coli* se réassocie avec un $Cot_{1/2}$ de 4,0. En comparant ces deux valeurs, nous déduisons de l'équation 5 que la complexité de la fraction est de $3,0 \times 10^8$ pb. Si l'on considère de la même façon les autres composants, l'on trouve une complexité de 6×10^5 pour le composant intermédiaire et une complexité de 340 pb seulement en ce qui concerne le composant rapide. Cela apporte une base quantitative à la relation décrivant que plus un composant se réassocie vite, plus sa complexité est faible.

A l'inverse, en prenant trois préparations d'ADN, chacune contenant une séquence unique de longueur appropriée (respectivement de 340 pb, 6×10^5 pb et

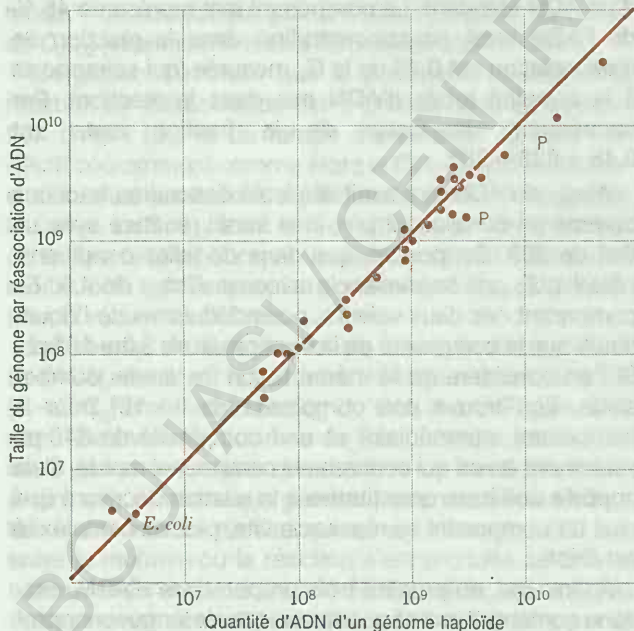
3×10^8 pb), et en les mélangeant dans les proportions de 25 :30 :45, chacune devrait se renaturer comme si elle était le composant unique décrit dans l'équation 2 ; l'ensemble des trois composants devrait alors décrire les mêmes cinétiques que celles déterminées par le génome total de la Figure 17.3.

LA COMPLEXITÉ DE L'ADN NON RÉPÉTITIF PEUT NOUS PERMETTRE D'ESTIMER LA TAILLE DU GÉNOME

La complexité du composant lent correspond à sa taille physique. Supposons que le génome se réassociant sur la Figure 17.3 ait une quantité d'ADN haploïde chimiquement déterminée de $7,0 \times 10^8$ pb. 45 % du génome représentent $3,15 \times 10^8$ pb, ce qui est à peine plus élevé que la valeur de $3,0 \times 10^8$ pb déduite des cinétiques de réassociation. Si l'on tient compte des erreurs de mesures inhérentes aux deux techniques utilisées, nous pouvons dire que la complexité du composant lent est la même quand elle est mesurée chimiquement et cinétiquement. Ces deux valeurs sont appelées respectivement la **complexité chimique** et la **complexité cinétique**.

Figure 17.4

Il existe une bonne corrélation entre la complexité de la cinétique de réassociation et la complexité des génomes eucaryotes, à l'exception des génomes polypléides (indiqués par P).



Le fait que ces deux valeurs coïncident signifie que le composant lent est constitué de séquences uniques dans le génome : après dénaturation, *chaque séquence monocaténaire ne peut se renaturer qu'avec la séquence complémentaire correspondante*. Cette fraction du génome est le seul composant de l'ADN procaryote et est généralement le composant majoritaire de l'ADN des eucaryotes. Il est appelé l'**ADN non répétitif**.

On peut utiliser la complexité cinétique de l'ADN non répétitif afin d'estimer la complexité du génome. Cela inverse simplement le calcul effectué précédemment. Dans l'exemple de la Figure 17.3, la complexité de l'ADN non répétitif est de $3,0 \times 10^8$ pb. Si cette fraction est unique et représente 45 % du génome, la totalité du génome devrait avoir une taille de $3,0 \times 10^8 : 0,45 = 6,6 \times 10^8$ pb. Cette estimation peut être comparée à la complexité chimique qui, dans ce cas, est de $7,0 \times 10^8$ pb.

La Figure 17.4 rend compte de la relation qui existe entre la taille du génome, déterminée par les cinétiques de réassociation d'ADN non répétitif, et la quantité haploïde d'ADN déterminée par analyse chimique. La correspondance entre ces deux valeurs démontre que le composant d'ADN non répétitif est constitué de séquences présentes en une seule copie dans le génome. Les seules exceptions sont les génomes de certaines plantes qui ont été engendrées par une récente polyploïdisation, et où, par conséquent, il y a plus d'une copie de chaque séquence.

Ceci marque un point important en ce qui concerne le paradoxe de la valeur C. En effet, la présence d'ADN non répétitif démontre que *les génomes de grande taille ne sont pas simplement le résultat d'une augmentation du nombre de copies des séquences présentes dans des génomes plus petits*. Si c'était le cas, les génomes de grande taille pourraient être considérés comme des génomes polypléides et alors il n'y aurait plus d'ADN non répétitif.

Les données obtenues grâce à la réaction de réassociation nous permettent d'exclure les modèles expliquant le paradoxe de la valeur C en terme de simple augmentation du nombre de copies de chaque gène présent dans un génome haploïde. Par conséquent, il faut raisonner sur la base suivante : plus un génome est important, plus sa diversité de séquences est grande.

LES GÉNOMES EUCARYOTES CONTIENNENT DES SÉQUENCES RÉPÉTITIVES

Quelle est la nature des composants qui se renaturent plus rapidement que l'ADN (lent) non répétitif ? Dans

l'exemple de la Figure 17.3, le composant intermédiaire occupe 30 % du génome. Sa complexité chimique est de $0,3 \times 7 \times 10^8 = 2,1 \times 10^8$ pb alors que sa complexité cinétique est seulement de 6×10^5 pb.

La longueur d'ADN qui correspond au $Cot_{1/2}$ de réassociation est beaucoup plus courte que la longueur totale de l'ADN chimiquement occupée par ce composant dans le génome. En d'autres termes, le composant intermédiaire se comporte comme s'il était constitué d'une séquence de 6×10^5 pb qui serait présente en 350 copies dans chaque génome (car $350 \times 6 \times 10^5 = 2,1 \times 10^8$). Après dénaturation, les brins monocaténaires issus de ces copies sont capables de se renaturer avec les brins complémentaires issus de n'importe laquelle des 350 copies. Par conséquent, les séquences entrant dans la réaction de réassociation sont en concentration plus élevée, justifiant le fait que le composant intermédiaire se renature à un $Cot_{1/2}$ plus bas.

Les séquences présentes à plus d'une copie dans chaque génome sont décrites comme étant l'**ADN répétitif**. Le nombre de copies présentes par génome est appelé la **fréquence de répétition (f)**. Formellement, la fréquence de répétition est définie par le rapport suivant :

$$f = \frac{\text{Complexité chimique}}{\text{Complexité cinétique}} \quad (6)$$

La complexité et la fréquence de répétition caractérisent les propriétés des composants du génome. Par exemple, si un génome donné est constitué d'une séquence longue de A pb, de 10 copies d'une séquence de B pb et de 100 copies d'une séquence de C pb, sa complexité est de $A + B + C$. Les fréquences de répétition des séquences A , B et C sont respectivement de 1, 10 et 100. La séquence A est non répétitive ; les séquences B et C sont répétitives.

La fraction d'ADN répétitif comprend tous les composants dont la fréquence de répétition est significativement supérieure à 1. Pour des raisons pratiques, les fréquences de répétition comprises entre 1 et 2 sont considérées comme correspondant à des séquences qui appartiennent au composant non répétitif (une séquence ne peut pas être présente autrement qu'un nombre entier de fois dans un génome haploïde !). Si l'on prend l'exemple de la Figure 17.3, l'équation 6 appliquée à l'ADN non répétitif donne une valeur de $3,15 \times 10^8 : 3,0 \times 10^8 = 1,05$. Généralement, les valeurs mesurées pour de l'ADN non répétitif dans des génomes réels sont d'environ 0,8-1,6.

Plutôt que de procéder à la série de calculs décrits précédemment, la fréquence de répétition de tout compo-

sant dans une courbe de réassociation peut être directement déterminée par comparaison avec le composant d'ADN non répétitif. La relation entre les fréquences de répétition de deux composants dans la même courbe de réassociation est donnée par le rapport inverse de leur $Cot_{1/2}$. Si l'on suppose que l'ADN non répétitif est vraiment unique (si f est égal à 1,0), nous pouvons poser une autre équation pour calculer la fréquence de tout composant répétitif :

$$f = \frac{Cot_{1/2} \text{ du composant d'ADN non répétitif}}{Cot_{1/2} \text{ du composant d'ADN répétitif}} \quad (7)$$

L'ADN MODÉRÉMENT RÉPÉTITIF CONSISTE EN DE NOMBREUSES SÉQUENCES DIFFÉRENTES

La fréquence de répétition est inversement proportionnelle au $Cot_{1/2}$. L'ADN répétitif est souvent classé en deux types, correspondant approximativement aux composants intermédiaire et rapide de la Figure 17.3. La fraction intermédiaire, se réassociant généralement à un Cot entre 10^{-2} et le Cot de l'ADN non répétitif, consiste en un **ADN modérément répétitif**. La fraction rapide, se réassociant avant que la valeur de Cot ait atteint 10^{-2} , consiste en un **ADN hautement répétitif**.

Le comportement d'un composant d'ADN répétitif ne représente qu'une moyenne utile à la description de ses séquences. Les paramètres décrivant ce composant ne sont pas nécessairement représentatifs des propriétés d'une séquence particulière issue de ce composant.

Le composant modérément répétitif de la Figure 17.3 est constitué d'une longueur totale d'ADN de 6×10^5 pb, répétée environ 350 fois par génome. Cela ne correspond pas à une longueur d'ADN continue et identifiable. *En réalité, cette longueur est composée d'une variété de séquences individuelles, toutes bien plus courtes*, qui mises bout à bout donnent une longueur totale de 6×10^5 pb. Ces séquences individuelles peuvent être dispersées tout le long du génome. Leur moyenne de répétition est de 350, certaines seront présentes en plus grand nombre alors que d'autres le seront en moins grand nombre.

Dans l'analyse d'un génome eucaryote, les composants sont rarement aussi bien séparés que sur la Figure 17.3. Bien souvent, ils se chevauchent tellement qu'ils sont indiscernables à l'œil. Un exemple est dessiné sur la Figure 17.5. Le génome de l'oursin de mer *S. purpuratus* se renature sur 8 ordres de grandeur de Cot ; les points d'inflexion de la courbe permettent de distinguer

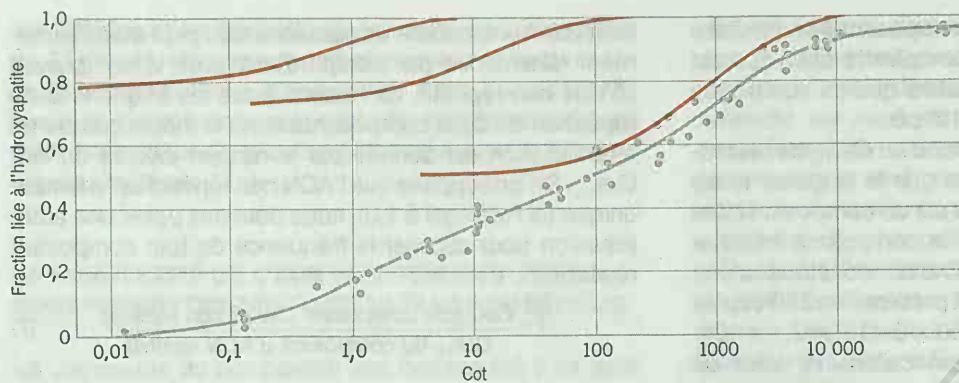


Figure 17.5

On peut discerner plusieurs composants dans les cinétiques de réassociation de l'ADN de *S. purpuratus*.

La courbe expérimentale est dessinée grâce à des points de données actuelles ; les courbes sans points de données expérimentales représentent les courbes idéales des composants qui pourraient rendre compte de la courbe expérimentale.

les différents composants mais la proportion et le $Cot_{1/2}$ de chacun d'entre eux n'est pas aisément calculable.

Afin d'identifier un à un les composants, un programme ordinateur qui possède toutes les courbes individuelles (courbes qui ont été dessinées pour d'hypothétiques composants de cinétique pure) s'accordant à chaque région de la courbe de réassociation est utilisé. Les courbes se chevauchent entre elles pour donner une courbe globale qui correspond à celle qui est observée pour le génome. Comme nous l'avons déjà mentionné, si des préparations d'ADN de complexité correspondante sont mélangées en certaines proportions, elle devraient donner lieu à la même courbe que celle obtenue expérimentalement.

Un exemple d'une telle analyse est donné sur la Figure 17.5. Les trois courbes correspondent à des composants occupant respectivement 19 %, 27 % et 50 % de l'ADN total (si l'on considère l'ordre de réassociation). Dans ce cas, les deux premiers composants peuvent être classés en tant qu'ADN modérément répétitif, avec des valeurs de $Cot_{1/2}$ de 0,53 et 8,3 ; la complexité de ces composants est respectivement de $1,0 \times 10^6$ pb répétée 160 fois et de $2,3 \times 10^6$ répétée 10 fois. Le troisième composant de la courbe est un ADN non répétitif.

Cette approche est généralement utilisée pour rechercher le nombre minimum de composants de cinétique différente dont la somme correspond de près à la courbe expérimentale. Cela ne reflète pas nécessairement la constitution biologique du génome. En effet, il est souvent possible d'obtenir une aussi bonne résolution avec un nombre plus important de composants. Par exemple, les

données de la Figure 17.5 peuvent décrire tout aussi bien la présence de trois composants modérément répétitifs occupant 10 %, 27 % et 25 % du génome avec des fréquences de répétition de 8 000, 240 et 19. Cette analyse réduit alors la taille du composant d'ADN non répétitif.

En réalité, il existe probablement un continuum de composants répétitifs, se réassociant de 10 fois à 20 000 fois plus rapidement que le composant d'ADN non répétitif. Les composants cinétiques que l'on a définis pour résoudre la courbe ne correspondent pas à des fractions répétitives discontinues, mais à des approximations utiles à des fins descriptives.

La proportion du génome occupée par de l'ADN répétitif est très variable. Dans le cas des eucaryotes inférieurs, la plupart de l'ADN est non répétitif, seulement 10-20 % de l'ADN appartient à un ou plusieurs composants modérément répétitifs. Dans les cellules animales, les composants modérément et hautement répétitifs occupent jusqu'à la moitié du génome. Chez les plantes et les amphibiens, la proportion d'ADN répétitif peut être encore plus importante et représenter la majeure partie du génome avec des composants modérément et hautement répétitifs représentant jusqu'à 80 % du génome. La longueur du composant d'ADN non répétitif tend à augmenter avec la taille du génome, mais elle ne lui est pas réellement proportionnelle.

Cette constatation vient modestement au secours du paradoxe de la valeur C si l'on considère que la variation des complexités de l'ADN non répétitif est moins importante que la variation des tailles de génome. Si l'on

suppose que les gènes font partie de l'ADN non répétitif, il reste à expliquer la raison pour laquelle la variation des complexités n'est pas proportionnelle aux caractéristiques morphologiques des différents organismes. Par exemple, la complexité d'ADN non répétitif de la mouche du vinaigre est de 1×10^8 pb alors que celle de la mouche domestique est de $2,8 \times 10^8$ pb ; il existe donc seulement un facteur trois entre leur complexité alors que ce facteur est de 6 si l'on compare leur taille génomique. La complexité de l'ADN non répétitif des mammifères est d'environ 9×10^8 pb alors que celle de la grenouille *X. laevis* est de $1,6 \times 10^9$ pb et celle des plantes et des amphibiens d'environ 9×10^9 pb.

En conclusion, nous pouvons dire qu'aucune caractéristique n'est propre aux génomes eucaryotes. Nous avons déjà mentionné que les plantes récemment devenues polyploïdes ne possèdent pas d'ADN non répétitif et que, par conséquent, la fraction de leur génome se réassociant le plus lentement a une fréquence de répétition de 2-3 ou plus. (En réalité, ceci représente probablement une étape intermédiaire de l'évolution, aboutissant à une divergence permettant de régénérer de l'ADN non répétitif.) Dans les génomes de crabes, il n'y a probablement pas d'ADN modérément répétitif mais simplement de l'ADN hautement répétitif et non répétitif. Chez les eucaryotes inférieurs, il ne peut pas y avoir d'ADN hautement répétitif.

LES MEMBRES DES FAMILLES DE SÉQUENCES RÉPÉTITIVES SE RESSEMBLENT MAIS NE SONT PAS TOUT À FAIT IDENTIQUES

Il est possible de séparer les différents composants de l'ADN eucaryote en arrêtant la réaction de renaturation à une certaine valeur de Cot. La **Figure 17.6** illustre cette expérience. Après que la réaction se soit produite à la valeur de Cot désirée, l'ADN renaturé est séparé en vertu de sa structure bicaténaires ; le matériel non réassocié reste monocaténaire. Afin de séparer les molécules bicaténaires des molécules monocaténaires, on utilise généralement des colonnes d'hydroxyapatite, qui ont la propriété de retenir les molécules d'ADN double brin.

La fraction bicaténaires correspond au matériel se réassociant rapidement (c'est-à-dire les séquences répétitives) ; la fraction monocaténaire correspond au matériel se réassociant plus lentement (c'est-à-dire l'ADN non répétitif et, en fonction de la valeur de Cot choisie, certaines séquences répétitives).

Une fois renaturé, l'ADN non répétitif est constitué de molécules bicaténaires qui se comportent de la même

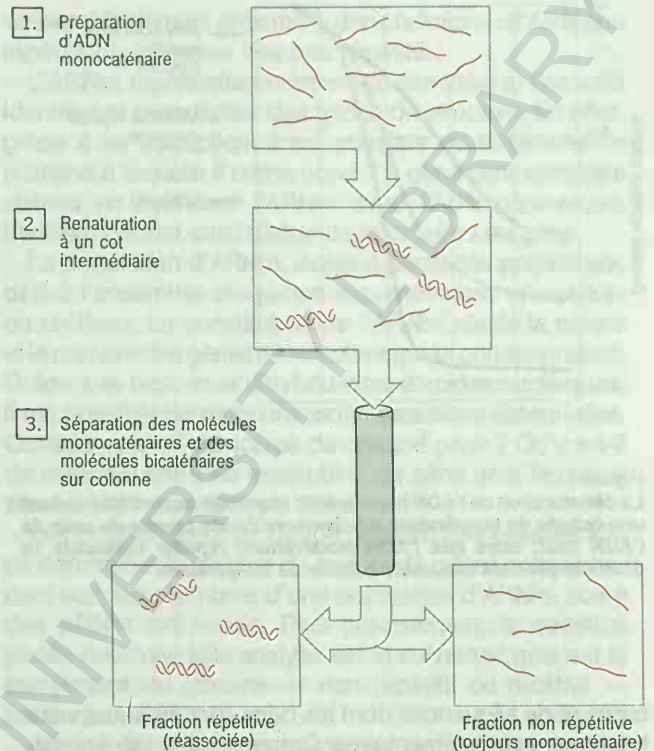


Figure 17.6

Les composants du génome de cinétique différente peuvent être isolés les uns des autres à un niveau intermédiaire de Cot.

façon que des molécules issues de la préparation d'ADN avant qu'elle soit dénaturée. Si on les dénature une seconde fois, les molécules bicaténaires se dissocient à un T_m seulement légèrement inférieur à celui de l'ADN natif initial. Ceci démontre que la réassociation des brins a bien eu lieu : chaque séquence unique s'est associée avec une séquence qui lui est strictement complémentaire.

L'ADN répétitif a un comportement tout différent quand il est renaturé. En effet, comme le montre la **Figure 17.7**, les molécules bicaténaires réassociées se séparent, non pas à une certaine température, mais à une échelle de températures assez large, ce qui signifie qu'elles ne sont pas constituées de brins strictement complémentaires. Plus il y a de mauvais appariements dans une molécule donnée, moins il y a de liaisons hydrogènes à rompre dans la réaction de dénaturation et, par conséquent, plus le T_m est bas.

L'amplitude de la courbe de dénaturation montre que l'ADN répétitif renaturé est constitué d'un éventail de molécules bicaténaires, c'est-à-dire de séquences issues de l'appariement de deux brins partiellement complémen-

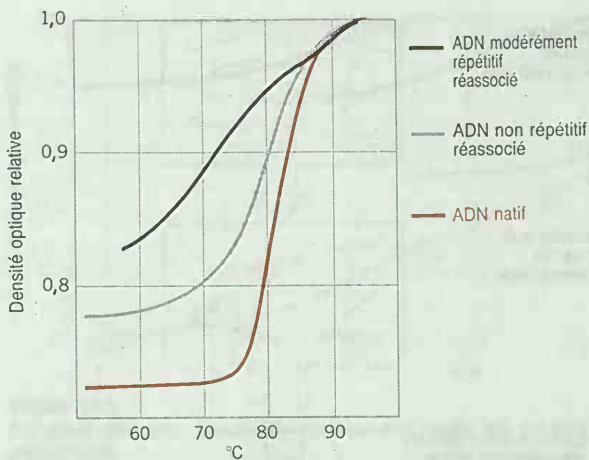


Figure 17.7

La dénaturation de l'ADN non répétitif réassocié se produit pendant une échelle de température relativement courte proche de celle de l'ADN natif, alors que l'ADN modérément répétitif réassocié se dissocie pendant une vaste échelle de température.

taires et de séquences dont les brins sont presque, voire totalement complémentaires. Comment cela est-il possible ?

Les composants d'ADN répétitif sont en réalité des familles de séquences qui ne sont pas tout à fait identiques mais se ressemblent. Les membres de chacune de ces familles sont constitués d'une série de séquences nucléotidiques qui se ressemblent suffisamment pour s'associer ensemble. Les différences qui existent entre membres d'une même famille sont le résultat de substitutions de bases, d'insertions et de délétions ; bien sûr, à ces endroits, il ne peut y avoir de complémentarité d'appariement. La proportion de ces événements permet d'établir une relation entre tout couple de séquences. En effet, quand deux membres d'une même famille très proche l'un de l'autre se renaturent ensemble, ils forment un double brin avec un T_m élevé, alors que quand deux membres de cette même famille mais de moindre ressemblance s'associent entre eux, ils forment un double brin avec un T_m plus bas.

Il est possible de contrôler la réassociation de molécules qui ne sont pas strictement identiques en faisant varier la **stringence** des conditions imposées lors de la réassociation. Une plus forte stringence est donnée en augmentant la température (par exemple), ce qui augmente le nombre d'appariements nécessaires pour obtenir une longueur d'ADN bicaténaire donnée. En réalisant les

réactions d'hybridation à température élevée, la réassociation peut être restreinte aux membres de grande ressemblance d'une famille donnée ; à température moins élevée, des membres moins ressemblants pourront s'associer ensemble. Notons que la proportion d'ADN génomique dans la fraction non répétitive augmente en même temps que la stringence.

Il existe une grande variation dans la constitution de familles répétitives. Certaines familles sont bien définies et sont constituées de membres de grande ressemblance. Elles restent intactes même quand la stringence d'hybridation augmente. En revanche, d'autres sont constituées d'un continuum de séquences de ressemblances variées, dont l'appartenance à cette famille devient de moins en moins évidente alors que la stringence augmente. La taille de telles familles répétitives est arbitraire puisqu'elle est déterminée en fonction des conditions d'hybridation.

Les génomes eucaryotes contiennent donc des séquences qui se ressemblent les unes aux autres, et qui le plus probablement sont issues d'une séquence ancestrale commune qui s'est multipliée, les éléments fils ayant divergé ; l'importance de cette divergence varie énormément.

La réassociation de deux séquences ressemblantes mais non identiques se produit plus lentement que la même réaction entre deux séquences strictement complémentaires. Le Cot nécessaire à la réassociation de séquences partiellement homologues est plus élevé que celui permettant l'appariement de deux brins totalement complémentaires, la valeur de $Cot_{1/2}$ observée pour des composants d'ADN répétitif peut ainsi être plus élevée que ne le voudrait la complexité réelle de ces fractions.

Les composants d'ADN répétitif ont par conséquent souvent des complexités moins importantes et des fréquences de répétition plus élevées que ne le laissent supposer leurs cinétiques de réassociation. Par exemple, des séquences ressemblantes qui ont une fréquence de répétition de 3-4 auront probablement des valeurs de $Cot_{1/2}$ qui les placeront dans la fraction d'ADN non répétitif. Une telle divergence entre $Cot_{1/2}$ et C est observée spécialement dans le cas de l'ADN hautement répétitif (voir Chapitre 21).

Où réside l'ADN modérément répétitif dans le génome ? Dans la plupart des génomes, il est dispersé entre les séquences non répétitives. Dans de nombreux cas, cette dispersion est assez régulière à travers le génome.

Le type d'organisation le plus courant est l'**espace-temps de période courte**, où l'ADN modérément répétitif est constitué de séquences d'une longueur moyenne d'environ 300 pb. Elles sont placées en alternance avec

des séquences non répétitives dont la longueur moyenne varie entre 800 et 1 500 pb. Ce type de modèle peut occuper la moitié du génome, et correspond à la plupart de l'ADN modérément répétitif et de l'ADN non répétitif.

Dans certains génomes on observe un **espacement de période longue** : il y a toujours une alternance des composants d'ADN modérément répétitif et d'ADN non répétitif, mais les séquences de chaque type sont beaucoup plus longues. Dans aucun des deux types d'espacements, la signification d'un tel arrangement n'a été élucidée.

LA PLUPART DES GÈNES DE STRUCTURE FONT PARTIE DE L'ADN NON RÉPÉTITIF

Par simplification, la génétique mendélienne sous-entend qu'il n'y a qu'une seule copie de chaque gène codant pour un facteur dans un génome haploïde. Chaque gène peut être localisé à une certaine place dans le génome, et l'hypothèse la plus simple est de supposer que chaque locus est occupé par une séquence d'ADN correspondant à une seule protéine, comme le met en relief la définition du gène dans le Chapitre 1.

La vision classique d'un gène de structure est donc la suivante : c'est un composant unique du génome, la seule séquence codant pour son produit protéique et par conséquent identifiable à l'aide de mutations qui font disparaître la fonction protéique qui lui est associée. La séquence d'un gène de structure unique, sans lien de parenté avec d'autres séquences du génome, doit donc faire partie du composant d'ADN non répétitif.

Dans le cas où des séquences multiples codent pour la même protéine, il est difficile voire impossible d'identifier chacun des gènes par mutation ponctuelle car l'inactivation d'une des copies n'empêche en rien l'expression des copies intactes. Les données génétiques sont biaisées car elles nous permettent uniquement de caractériser des gènes uniques. Il faut analyser directement l'ADN du génome pour déterminer les nombres et proportions de séquences uniques et répétées.

Toutefois, un gène de structure peut être unique parce qu'il est le seul à coder pour une certaine protéine, mais il peut aussi ressembler à d'autres séquences du génome qui codent pour des protéines apparentées. En outre, à cause de la dégénérescence du code génétique, les séquences qui codent pour le même polypeptide ne sont pas nécessairement identiques. Généralement, une famille de ce type n'est constituée que de quelques membres ; s'il y a de mauvais appariements après la réaction de réassociation, les gènes de cette famille vont

vraisemblablement apparaître dans la fraction d'ADN non répétitif ou du moins très peu répétitif.

L'ARNm représente un intermédiaire idéal si l'on veut identifier et caractériser des gènes de structure. En effet, grâce à sa traduction, il est possible de déterminer la protéine à laquelle il correspond. Le gène peut alors être obtenu en hybridant l'ARNm avec l'ADN génomique. L'ARNm permet ainsi d'aller de la protéine au gène.

La population d'ARNm, isolée à partir des polysomes, définit l'ensemble des gènes exprimés dans une cellule ou un tissu. La constitution de l'ARNm révèle la nature et le nombre des gènes de structure qui lui correspondent. Grâce à la technique d'hybridation d'acides nucléiques, il est possible de répondre à des questions essentielles. Combien y a-t-il de copies de chaque gène ? Qu'y a-t-il de commun entre les ensembles de gène dont l'expression définit tel ou tel type cellulaire ?

Les cinétiques d'hybridation peuvent être utilisées afin de déterminer le nombre de copies de gènes correspondant soit aux membres d'une population d'ARNm, soit à des ARNm individuels. Plus précisément, la question posée pour une telle analyse est la suivante : quel est le composant du génome — non répétitif ou répétitif — transcrit en ARNm ?

Afin de déterminer les composants du génome représentés dans les ARNm, l'ARN est utilisé comme **traceur** dans la réaction de réassociation. Une très faible quantité d'ARN (ou d'ADN) marqué radioactivement est ajoutée à une quantité beaucoup plus importante d'ADN cellulaire. L'ADN cellulaire étant présent en large excès, la concentration de séquences d'ADN monocaténares reste inchangée après hybridation de certaines séquences avec l'ARN traceur.

La réaction permet la réassociation des brins d'ADN cellulaire complémentaires (comme si l'ADN traceur n'était pas présent). Ceci est décrit comme étant une **réaction dirigée par l'ADN**. La réassociation de la totalité de l'ADN est suivie grâce aux moyens habituels (modification de la densité optique ou rétention sur colonne d'hydroxyapatite). L'entrée de la radioactivité dans la fraction bicaténaire permet de suivre la réassociation du traceur.

L'ARN traceur (ou ADN) participe à la réaction comme s'il était membre du composant à partir duquel il a été transcrit. Le composant est alors identifié en comparant la courbe obtenue avec le traceur radioactif et la courbe de réassociation de la totalité de l'ADN. Les valeurs de *Cot* pour lesquelles l'ARN marqué s'hybride permettent de définir les fréquences de répétition des séquences génomiques correspondantes.

Lorsqu'une seule espèce de molécule d'ARN est utilisée dans la réaction, elle s'hybride à une seule valeur de $Cot_{1/2}$ déterminée par la fréquence de répétition du gène ou des gènes lui correspondant. Si le gène est unique, cette valeur sera la même que celle d'un ADN non répétitif ; si le gène est répété, la valeur du $Cot_{1/2}$ diminuera en fonction de la fréquence de répétition. Lorsque toute une population d'ARN est utilisée dans la réaction, chaque espèce d'ARN hybride à un $Cot_{1/2}$ particulier ; la courbe globale correspond à la somme des courbes qui seraient obtenues pour chacun des composants. Elle peut être résolue de la même façon que l'est la courbe de réassociation de l'ADN génomique.

Si l'on utilise une population d'ARNm, le résultat obtenu ressemble généralement à celui qui est décrit sur la **Figure 17.8** Une faible proportion d'ARN, 10 % au moins, s'hybride à un $Cot_{1/2}$ correspondant à des séquences modérément répétitives. Le composant majeur, lui, s'hybride avec un $Cot_{1/2}$ proche ou identique à celui de l'ADN non répétitif. Généralement, il représente jusqu'à 50 % de l'ARN total.

Quelle relation existe-t-il entre les séquences d'ARNm qui s'hybrident avec l'ADN non répétitif et celles qui s'hybrident avec l'ADN répétitif ? Les ARNm peuvent être séparés en deux catégories (en récupérant l'ARN qui s'hybride le premier), ce qui signifie que des types différents de molécules sont impliqués dans la réaction. Ces deux catégories sont, d'une part, les gènes qui font

partie de l'ADN non répétitif et, d'autre part, les gènes qui font partie de l'ADN répétitif.

Le problème majeur dans l'interprétation des résultats est de décider de la nature des séquences qui ne s'hybrident pas du tout, séquences qui représentent jusqu'à la moitié de l'ARN total. La plupart de ces séquences dérivent probablement des séquences non répétitives. En effet, une raison expliquant que ces ARN ne s'hybrident pas pourrait être qu'ils sont présents en très grande quantité rendant alors inexacte l'approximation selon laquelle la quantité d'ARN dans le traceur est petite par rapport à celle de l'ADN auquel il peut s'hybrider. En d'autres termes, si l'excès d'ADN n'est pas suffisant dans une réaction dirigée par l'ADN, certains ARN ne s'hybrideront pas.

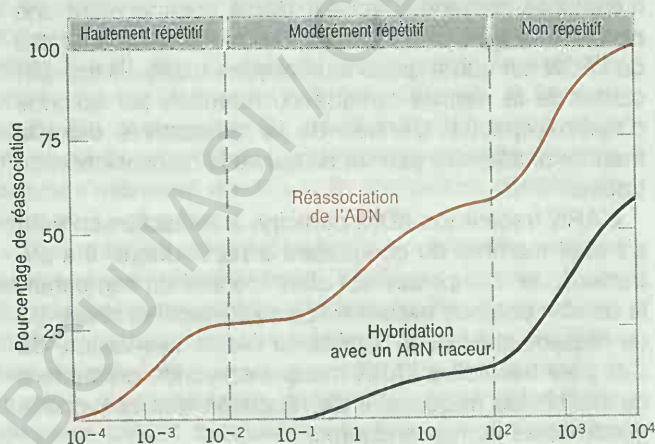
Au vu de ces résultats, il devient clair que la majorité de l'ARNm, peut-être jusqu'à 80 %, dérive des séquences se réassociant avec le composant d'ADN non répétitif. Puisqu'il est difficile de détecter de très faibles degrés de répétition (spécialement, quand les copies répétées se ressemblent ou sont identiques), ces gènes ne sont pas nécessairement uniques, mais peuvent être présents à moins de trois ou quatre copies par génome. Seule une faible proportion de gènes fait de toute évidence partie de l'ADN répétitif ; les cinétiques d'hybridation ne révèlent pas si ces multiples copies sont identiques ou seulement ressemblantes.

Cependant, bien que la majorité des gènes semblent être inclus dans le composant d'ADN non répétitif, cette technique ne prouve pas que 80 % d'entre eux en font partie car les expériences décrites sont menées en utilisant la *masse totale d'ARN* et les niveaux d'expression des gènes sont différents.

Les cinétiques d'hybridation avec un seul ARNm ont été déterminées pour plusieurs gènes. Généralement, les résultats suggèrent qu'il y a une ou deux copies du gène. En réalité, cette technique sous-estime le nombre de séquences ressemblantes dans le génome, puisque les expériences permettant d'isoler directement un gène identifient aussi d'autres séquences lui ressemblant (voir Chapitre 19).

Figure 17.8

La courbe de réassociation donnée par hybridation d'une préparation d'ARNm traceurs montre que parmi les séquences d'ARNm, une majorité dérive de l'ADN non répétitif, une minorité de l'ADN modérément répétitif et aucun de l'ADN hautement répétitif.



COMBIEN Y A-T-IL DE GÈNES NON RÉPÉTITIFS EXPRIMÉS ?

Le nombre de séquences d'ADN non répétitif transcrites en ARN peut être directement déterminé grâce à la proportion d'ADN pouvant s'hybrider avec l'ARN. Lorsqu'une faible quantité d'ADN simple brin est hybridée

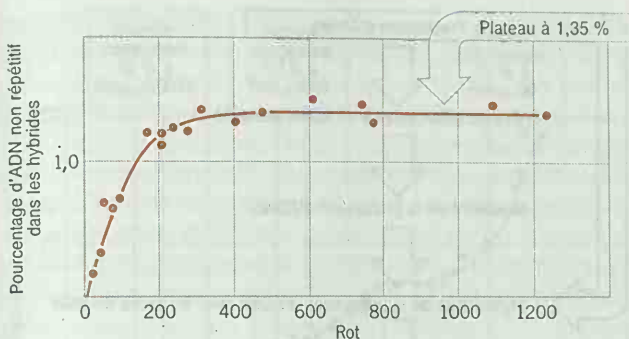


Figure 17.9

L'hybridation d'un excès d'ARNm avec de l'ADN non répétitif, jusqu'à ce qu'il y ait saturation, montre que seule une petite proportion de l'ADN est représenté en ARNm.

L'expérience est réalisée en utilisant de l'ADN non répétitif purifié qui est marqué radioactivement ; ainsi, l'on peut mesurer la quantité de radioactivité entrée en hybridation. Les données expérimentales représentées sur la figure proviennent de la réaction d'hybridation entre l'ADN non répétitif de l'oursin de mer *S. purpuratus* et de l'ARNm extrait des polysomes de l'embryon au stade gastrula.

avec une grande quantité d'ARN, toutes les séquences de l'ADN complémentaires à l'ARN peuvent réagir pour former un hybride ARN-ADN.

Dans ce type d'expérience à saturation, il faut un excès d'ARN suffisamment important pour que chaque séquence d'ADN complémentaire soit hybridée. Comme la réaction est dirigée par l'ARN, le paramètre de contrôle est le produit de la concentration d'ARN avec le temps, paramètre que l'on appelle le **Rot**. Il est tout à fait comparable au **Cot** décrit dans les réactions dirigées par l'ADN.

Le tracé du pourcentage de l'ADN hybridant en fonction de la valeur de Rot permet de suivre l'hybridation à saturation. Une telle courbe est décrite sur la **Figure 17.9**. Dans cet exemple, la réaction est totale à une valeur de Rot d'environ 300, mais il est nécessaire de l'augmenter pour s'assurer que le plateau a bien été atteint. Quand la valeur de rot atteint 1 200, plus aucune molécule d'ADN ne s'hybride, ce qui prouve que tout l'ADN non répétitif a bien été hybridé.

La proportion d'ADN hybridé permet de calculer le nombre de gènes représentés dans la population d'ARNm utilisé pour conduire la réaction. A saturation, 1,35 % de l'ADN non répétitif est hybridé. Puisque seul un des brins d'ADN est transcrit en ARN, seule la moitié de l'ADN est potentiellement capable de s'hybrider avec l'ARN (la seconde moitié a la même séquence que l'ARN). Par

conséquent, 2,70 % des séquences totales de l'ADN non répétitif sont représentés dans les ARNm.

L'ADN non répétitif représente 75 % d'un génome de $8,1 \times 10^8$ pb. La complexité de l'ADN représenté dans la population d'ARNm est donc de $0,027 \times 0,75 \times 8,1 \times 10^8 = 1,7 \times 10^7$ pb. La population d'ARNm a une longueur moyenne de 2 000 bases. Le nombre total de messagers différents est donc de $1,7 \times 10^7 / 2\,000 = 8\,500$. Cela correspond au nombre total de gènes exprimés dans le tissu d'où sont issus les ARNm (l'embryon au stade gastrula de l'oursin de mer).

Des expériences similaires ont été réalisées pour différents systèmes. Pour un eucaryote inférieur comme la levure, le nombre total de gènes exprimés est relativement bas, d'environ 4 000. Dans les tissus somatiques des eucaryotes supérieurs, ce nombre se situe généralement entre 10 000 et 15 000. Cette valeur est identique chez les plantes et les vertébrés. La quantité totale d'ADN représentée dans les ARNm correspond à une petite proportion du génome, de l'ordre de 1-2 %. La seule exception est illustrée par le cerveau de mammifère où un plus grand nombre de gènes sont exprimés, bien que la quantification exacte soit incertaine.

Ce type d'expérience ne peut être réalisée que pour de l'ADN non répétitif, dans lequel chaque séquence d'ADN ne peut s'hybrider qu'avec son ARNm. A cause de la présence de multiples copies de séquences identiques ou ressemblantes dans l'ADN modérément répétitif, un ARN issu de ce composant peut probablement s'hybrider avec d'autres séquences génomiques en plus de la séquence qui a permis sa transcription. A saturation, un grand nombre de séquences d'ADN supplémentaires peuvent faire partie des hybrides, ce qui a pour résultat une surestimation du nombre de gènes exprimés.

En fait, pour être certain que cela n'a pas eu lieu, l'ADN qui a été saturé par l'ARN est recouvré (en dégradant l'ARN), puis utilisé dans une expérience de réassociation afin de montrer que son $Cot_{1/2}$ correspond bien à l'ADN non répétitif attendu. C'est ce que l'on appelle une **expérience de « playback »**. L'estimation obtenue grâce à cette technique démontre que la majorité des séquences exprimées font partie de l'ADN non répétitif.

LES CINÉTIQUES DES RÉACTIONS DIRIGÉES PAR L'ARN PERMETTENT D'ESTIMER LE NOMBRE DE GÈNES

Puisque l'ARN est en excès par rapport à l'ADN, la petite quantité d'ARN qui s'hybride durant la réaction dirigée

par l'ARN ne modifie pas substantiellement la concentration d'ARN. Cela signifie qu'une analyse à saturation suit des cinétiques de premier ordre, comme le décrit l'équation suivante :

$$\frac{D}{D_0} = e^{-k \cdot Rot_{1/2}} \quad (8)$$

lorsque la réaction est accomplie à moitié et que $D/D_0 = 0,5$,

$$K = \frac{\ln 2}{Rot_{1/2}} \quad (9)$$

Le $Rot_{1/2}$ d'une réaction dirigée par l'ARN à saturation ne peut pas être utilisé pour déterminer la complexité d'une population d'ARN. En effet, les séquences d'ARN étant présentes à différentes concentrations définies par le taux d'expression du (des) gène(s) dont ils sont issus, le Rot mesuré pour une masse totale d'ARN ne peut être considéré comme celui de chacune des séquences prises individuellement.

En réalité, la majeure partie de la masse d'ARNm n'est souvent composée que de quelques séquences. Elles satureront très rapidement leur complément dans l'ADN, ce qui permet à la réaction d'être dirigée par les séquences restantes. Par conséquent, la masse réelle d'ARN dirigeant la réaction est bien inférieure au R_0 mesuré. Le $Rot_{1/2}$ mesuré est de ce fait beaucoup trop élevé car il rend compte de l'importante masse d'ARN dont les séquences complémentaires ont été rapidement saturées et qui, par conséquent, ne participe plus à la réaction.

Une autre technique a été développée afin de déterminer la complexité grâce aux cinétiques de la réaction d'hybridation dirigée par l'ARN. Le protocole de cette technique est illustré sur la **Figure 17.10**. Il utilise un ADNc marqué préparé par transcription inverse de l'ARNm. Rappelons qu'un ADNc est une molécule d'ADN monocaténaire complémentaire de l'ARNm.

La réaction consiste à hybrider l'excès d'ARNm avec des molécules d'ADNc marquées qui ont été préparées à partir de la totalité des ARN. Chaque séquence d'ARNm doit donc être représentée dans la population d'ADNc à une fréquence correspondant à sa proportion dans l'ARN total. Puisque toutes les séquences d'ADNc dérivent des ARNm, toutes les molécules d'ADNc marquées doivent hybrider. Pour un composant unique, le $Rot_{1/2}$ de la réaction est proportionnel à sa complexité de même que le $Cot_{1/2}$ est déterminé par la complexité de l'ADN dans les réactions de réassociation. Par conséquent, la com-

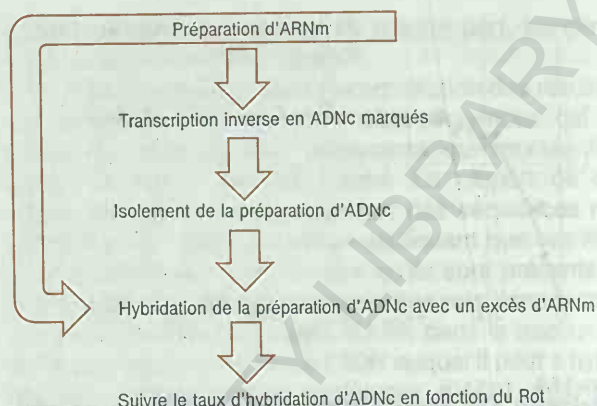


Figure 17.10

L'hybridation d'un excès d'ARNm avec des ADNc est utilisée pour déterminer la complexité des ARNm.

plexité d'une population d'ARN inconnue peut être déterminée en comparant son $Rot_{1/2}$ avec le $Rot_{1/2}$ d'une réaction standard.

La réaction entre un excès d'ARNm de globine et son ADNc est souvent utilisée comme standard. L'on peut donc écrire :

$$\frac{\text{Complexité de tout ARN}}{\text{Complexité de l'ARNm de globine}} = \frac{Rot_{1/2} \text{ de tout ARN}}{Rot_{1/2} \text{ de l'ARNm de globine}} \quad (10)$$

Cette équation est analogue à l'équation 5 qui détermine la complexité de préparations d'ADN inconnues en comparant leurs valeurs de $Cot_{1/2}$ à celle d'un ADN standard. Dans l'analyse des Rot , la complexité calculée rend également compte de la longueur totale de différentes séquences sans préjuger de l'organisation de ces séquences.

Tout comme dans la courbe de réassociation de l'ADN, un composant unique s'hybride sur deux ordres de grandeur de Rot , et une réaction qui s'étend sur une échelle plus large doit être résolue par ordinateur afin que soient définis individuellement les différents composants qui la constituent. Rappelons que la courbe peut tout aussi bien correspondre à un spectre continu de séquences.

Afin de déterminer leur complexité, on considère que chaque composant est le seul à être présent dans la réaction. La valeur de $Rot_{1/2}$ mesurée pour un composant est alors multipliée par la proportion qu'occupe ce composant dans la réaction. La valeur de $Rot_{1/2}$ ainsi corrigée est utilisée pour calculer la complexité du composant.

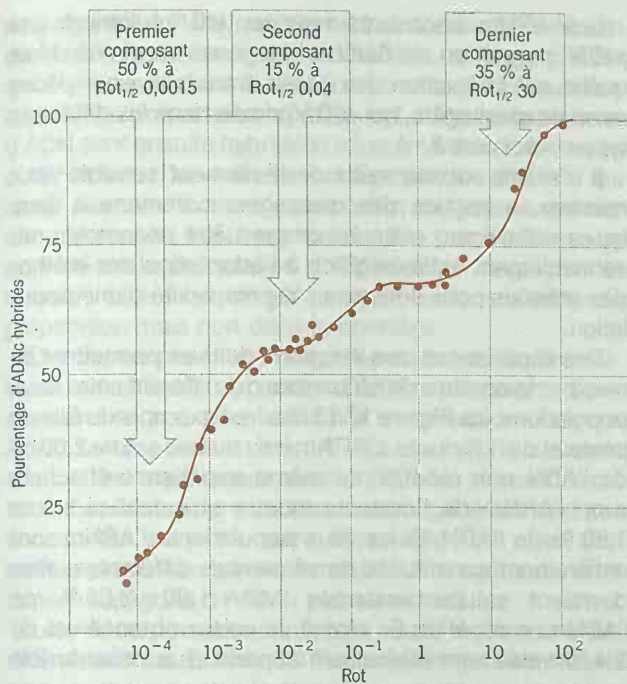


Figure 17.11

L'hybridation d'un excès d'ARNm et d'ADNc permet l'identification de plusieurs composants, chacun d'entre eux caractérisé par le Rot_{1/2} de la réaction.

La réaction est exprimée relativement à la quantité totale d'ADNc hybridée bien que cette valeur soit généralement proche de 100 %, presque toujours supérieure à 90 %. Les données de la figure sont issues d'une réaction utilisant les ARNm de l'oviducte de poule.

Un exemple de réaction excès d'ARN/ADNc engendrant trois composants est décrit sur la **Figure 17.11**. L'hybridation entre l'ARNm purifié de l'ovalbumine (long de 2 000 pb) et son ADNc, qui a un Rot_{1/2} de 0,0008 fut prise comme contrôle de la réaction. Cela signifie qu'il faut un Rot_{1/2} de 0,0004 pour une séquence dont la complexité est de 1 000 bases.

Le premier composant a une valeur de Rot_{1/2} observée de 0,0015 et représente 50 % de la totalité des composants. Par conséquent, la valeur du Rot_{1/2} de ce composant purifié doit être de $0,5 \times 0,0015 = 0,00075$, pratiquement identique à celle du Rot_{1/2} de l'ovalbumine prise comme contrôle, suggérant que le premier composant est justement l'ARNm de l'ovalbumine (qui occupe environ la moitié de la masse des messagers dans l'oviducte).

Le composant suivant a une valeur de Rot_{1/2} observée de 0,04, et représente 15 % de la totalité des composants. Par conséquent, son Rot_{1/2} est de

$0,15 \times 0,04 = 0,006$. Sa complexité est alors de $0,006 : 1\,000 \times 0,0004 = 15\,000$ bases. Cela doit correspondre à 7-8 espèces d'ARN d'une longueur moyenne de 2 000 bases.

Le dernier composant a un Rot_{1/2} de 30 et représente 35 % de la totalité des composants ; il est purifié à un Rot_{1/2} de $0,35 \times 30 = 10,5$ ce qui représente une complexité de $10,5 : 0,0004 \times 1\,000 = 26\,000\,000$ bases. Cela correspond à 13 000 espèces d'ARNm d'une longueur moyenne de 2 000 bases.

Cette analyse met en évidence qu'approximativement la moitié de la masse d'ARNm de la cellule correspond à une seule sorte d'ARNm, 15 % à un ensemble de 7-8 ARNm et 35 % à environ 13 000 espèces d'ARNm. Par conséquent, il est clair que les ARNm de chaque composant sont présents à des taux différents.

LES GÈNES SONT EXPRIMÉS À DES TAUX TRÈS VARIABLES

Pour chaque ARN, le nombre moyen de molécules par cellule est appelé l'**abondance** ou la **représentation**. Il peut être calculé assez simplement si la masse totale de l'ARN dans la cellule est connue. Pour chaque composant, la masse totale = abondance $\times$ complexité, ce qui nous permet d'écrire l'équation suivante :

$$\text{Abondance} = \frac{\text{gm d'ARNm dans la cellule} \times \text{fraction du composant} \times 6 \times 10^{23}}{\text{Complexité du composant en daltons}} \quad (11)$$

L'équation est généralement posée ainsi car l'ARNm total est mesuré (chimiquement) en picogrammes, alors que la complexité est exprimée (par hybridation) en bases ou en daltons.

Dans l'exemple de la Figure 17.11, il y a 0,275 pg d'ARNm par cellule. Cela correspond à 100 000 copies du premier composant (l'ARNm de l'ovalbumine), 4 000 copies de chacun des 7-8 ARNm du second composant, et seulement environ 5 copies de chacun des 13 000 ARNm du dernier composant.

L'oviducte, où une telle quantité d'ARNm ne représente qu'une seule espèce d'ARNm, est un cas extrême. La majorité des cellules contiennent un petit nombre d'ARNm, chacun étant présent en plusieurs copies. Le composant **abondant** consiste également en moins d'une centaine d'ARNm différents présents à plus de 1 000 voire de 10 000 copies par cellule. Il correspond souvent à la majeure partie de la masse d'ARN c'est-à-dire à presque la moitié de l'ARNm total. Le composant

abondant peut être séparé en deux composants, comme le montre l'exemple de l'oviducte, il peut également ne comporter qu'un seul composant ; mais en fait, le niveau de transcription des gènes codant pour les messagers abondants est très différent d'un cas à l'autre.

Dans tous les cas, approximativement la moitié de la masse d'ARNm consiste en un grand nombre de séquences, de l'ordre de 10 000, chacune représentée seulement en un petit nombre de copies dans l'ARNm total, inférieur à 10. C'est la classe d'ARNm **peu abondante** ou **complexe**. (C'est cette classe qui dirige une réaction de saturation.)

Le nombre total de gènes exprimés, estimé d'une part par la technique de saturation et d'autre part par la technique de cinétique, est généralement proche. La technique de cinétique donne une sous-évaluation (car certaines séquences très peu abondantes ne peuvent réagir). La technique de saturation donne une sur-évaluation (car il y a toujours des séquences qui sont présentes en plus d'une copie par génome). Par exemple, pour l'oviducte de poule, la technique cinétique permet l'identification d'environ 13 000 ARNm alors que la technique de saturation en définit environ 15 000.

QUELLES SONT LES PARTIES COMMUNES ENTRE LES DIFFÉRENTES POPULATIONS D'ARNm ?

Dans de nombreux tissus somatiques d'eucaryotes supérieurs, le nombre de gènes exprimés se situe entre 10 000 et 20 000. Quelle est la proportion de gènes qui s'expriment dans tous les tissus ? Par exemple, le nombre de gènes exprimés dans le foie de poule est d'environ 11 000-17 000, alors qu'il est d'environ 13 000-15 000 dans l'oviducte. Combien y a-t-il de gènes exprimés qui sont identiques entre ces deux tissus ? Combien y a-t-il de gènes exprimés qui sont spécifiques de chacun des tissus ?

Il est évident que les gènes exprimés appartenant à la catégorie abondante diffèrent selon les tissus. L'ovalbumine, par exemple, est uniquement synthétisée dans l'oviducte, et pas du tout dans le foie. Cela signifie que 50 % de la masse d'ARNm dans l'oviducte est spécifique de ce tissu. Dans de nombreux cas, les ARNm abondants correspondent à des gènes codant pour des protéines qui sont des produits majoritaires dans le type cellulaire considéré. En terme de masse, il peut y avoir des différences importantes entre les messagers et les protéines synthétisées par différents types cellulaires. A la limite, certains gènes ne sont exprimés que dans certaines cellules.

Les ARNm abondants représentent seulement une petite proportion du nombre de gènes exprimés. Mais quelle est la fraction des gènes transcrits en espèces rares de messagers, qui est exprimée dans les différents types cellulaires ?

Il n'existe aucune méthode réellement sensible pour mesurer la fraction des messagers commune à deux types cellulaires ; cette valeur peut être néanmoins raisonnablement estimée grâce à l'adaptation des méthodes utilisées pour déterminer la complexité d'une population.

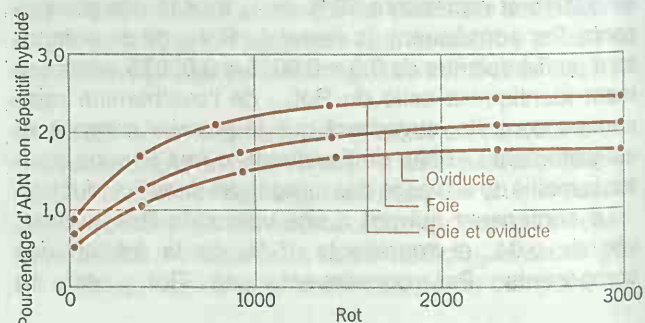
Des expériences de saturation additives permettent de mesurer le nombre de séquences qui diffèrent entre deux populations. La **Figure 17.12** illustre l'exemple du foie de poule et de l'oviducte. L'ARNm issu du foie sature 2,05 % de l'ADN non répétitif ; la même expérience effectuée avec l'ARNm de l'oviducte montre que celui-ci sature 1,80 % de l'ADN. Si les deux populations d'ARNm sont entièrement constituées de séquences différentes, elles devraient saturer ensemble $2,05 + 1,80 = 3,85$ % de l'ADN non répétitif. En réalité, la valeur obtenue est de 2,4 % seulement légèrement supérieure à celle du foie seul.

Par conséquent, environ 75 % des séquences exprimées dans le foie et l'oviducte sont les mêmes (l'expérience étant réalisée à saturation, les données ne permettent pas de savoir si ces séquences sont présentes à des taux identiques ou différents dans les deux tissus). En d'autres termes, environ 12 000 gènes sont exprimés à la fois dans le foie et l'oviducte, environ 5 000 autres gènes sont exprimés uniquement dans le foie, et environ 3 000 gènes sont exprimés uniquement dans l'oviducte.

Un autre moyen pour évaluer la partie commune d'ARNm est d'hybrider l'ARNm d'un tissu avec de l'ADN

Figure 17.12

Des hybridations additives à saturation permettent de calculer le nombre de séquences différentes présentes dans deux préparations d'ARNm.



non répétitif. L'ADN qui s'hybride est ensuite isolé afin de constituer une préparation d'ADNm, alors que l'ADN qui n'entre pas dans la réaction constituera ce que l'on appelle la préparation d'ADN nul. Ces préparations d'ADN sont ensuite hybridées séparément avec un excès d'ARNm issu d'un autre tissu. La proportion d'ADNm qui réagit permet d'évaluer la proportion de gènes exprimés dans le second tissu comme dans le premier tissu. La quantité de la préparation d'ADN nul qui réagit permet d'évaluer le nombre de gènes exprimés dans la seconde préparation mais non dans la première.

Des expériences hétérologues peuvent aussi être réalisées en utilisant un excès d'ARNm provenant d'une première source pour conduire une réaction avec un ADNc correspondant à un ARNm issu d'une seconde source. Ces expériences montrent que les séquences qui sont abondantes dans un tissu ne sont généralement pas abondantes dans un autre tissu, même si elles sont parfois présentes (généralement à taux moins élevé et dans la catégorie d'ARNm peu abondant).

Les ARNm peu abondants sont souvent communs aux différents tissus. Environ 90 % des ARNm peu abondants sont identiques entre le foie et le rein de souris, ce qui laisse seulement une différence de 1 000-2 000 en terme de gènes exprimés. Généralement, les résultats obtenus après de telles comparaisons montrent que seuls 10 % des ARNm d'une cellule en sont spécifiques. La majorité des séquences est commune à de nombreux, si ce n'est à tous les types cellulaires.

Cela suggère que dans l'ensemble des fonctions assurées par les gènes exprimés, fonctions qui sont environ au nombre de 10 000 chez un mammifère, il y a des fonctions qui sont nécessaires à tous les types cellulaires. Quelquefois, ce type de fonction est appelé activité **de ménage** ou **constitutive**. Elle s'oppose aux activités qui correspondent à des fonctions spécialisées (comme l'ovalbumine ou la globine) qui ne sont présentes que dans des types cellulaires particuliers. Celles-ci correspondent à ce que l'on appelle souvent les fonctions de **luxé**.

Si nous considérons tous les types cellulaires d'un organisme, il y a au moins autant de fonctions de luxe que de fonctions de ménage. Au plus, il est improbable que le nombre total de fonctions de luxe (du moins dans l'ADN non répétitif) dépasse, disons de 2-3 fois celui des

fonctions de ménage — soit dans la gamme 20 000-40 000.

Certaines cellules spécialisées peuvent fonctionner avec un nombre relativement peu élevé de gènes exprimés. Un examen détaillé des gènes exprimés dans l'oursin de mer *S. purpuratus* révèle que l'oocyte contient environ 18 000 séquences d'ARNm. Au cours de son embryogenèse, le nombre de gènes exprimés n'est plus que d'environ 13 000 au stade blastula, 8 500 au stade gastrula et 7 000 au stade pluteus. Mais dans certains des tissus adultes — tube digestif, intestin, cœlome — le nombre de gènes exprimés est seulement de 2 500-3 000. Ces cellules peuvent donc être maintenues avec un nombre de fonctions bien inférieur au nombre de fonctions de ménage qui est de 10 000 dans les cellules de mammifère.

Une caractéristique importante de cette série de cellules est qu'une importante partie des séquences qu'elles expriment leur est commune, et que leur développement génétique se déroule grâce à une réduction progressive du nombre de gènes exprimés. Les expériences de dosages d'ADNm/ADN nul ont montré que la majorité des gènes exprimés à un stade avancé ont aussi été exprimés à un stade précoce embryonnaire. Il est même possible que tous les gènes de structure d'un organisme soient exprimés dans l'oocyte, ce qui signifierait que le génome de l'oursin de mer est constitué de plus de 20 000 gènes dont les séquences codantes correspondent seulement à 3 % de l'ADN total haploïde. Même si l'on considère que chaque gène n'est pas unique mais répété disons 3 fois dans le génome, il reste encore une grande partie du génome qui ne semble pas coder pour des protéines.

POUR EN SAVOIR PLUS

La source principale d'où sont issues les informations discutées dans ce chapitre est *Gene expression* de Lewin, 2, *Eucaryotic Chromosomes* (Wiley, New York, 1980, pp. 479-530 et 694-727). La technique de réassociation fut décrite par Britten & Kohne (*Science*, 161, 529-540, 1968) et les résultats obtenus grâce à elle furent retracés par Britten & Davidson dans une revue (*Quart. Rev. Biol.* 48, 565-613, 1973). Une vision contemporaine de l'analyse du nombre de gènes fut résumée par Lewin (*Cell* 4, 77-93, 1975).

CHAPITRE 18

L'ORGANISATION DES GÈNES MORCELÉS

Pendant longtemps, il fut suspecté que les gènes eucaryotes étaient fondamentalement différents des gènes bactériens en raison des différences de quantités d'ADN que décrit le paradoxe de la valeur C. Les mesures de complexité des gènes de structure qui démontrent que seule une faible partie du génome est représentée dans l'ARNm, vinrent renforcer cette idée.

Afin d'expliquer ces résultats, le plus simple serait de supposer qu'une proportion importante (peut-être même la majorité) des séquences d'ADN ne fait pas partie des gènes de structure. La quantité et la fonction (s'il en a une) d'un tel ADN restent à définir.

Une autre hypothèse serait de supposer que la taille d'un gène de structure, ou du moins celle de l'unité de transcription, est bien plus importante que celle de l'ARNm. Indirectement, le fait que la taille de l'ARN nucléaire soit bien plus importante que celle de l'ARN est en accord avec cette hypothèse. La majeure partie de cette longueur « supplémentaire » est supprimée lors de la **maturation** de l'ARNm c'est-à-dire lorsque celui-ci est transporté dans le cytoplasme.

Afin d'expliquer cette augmentation de la taille du gène, on supposa tout d'abord qu'une partie de l'une ou des deux séquences adjacentes à la région correspondant à l'ARNm était incluse dans l'unité de transcription. Les régions adjacentes non transcrites (plus probablement celles se situant en amont de l'unité de transcription) pourraient être nécessaires à la régulation du gène ; de ce fait, la taille de l'unité permettant l'expression d'un gène peut être plus importante que celle de l'ARNm correspondant.

De nombreux gènes eucaryotes ont une taille bien plus importante que celle de leur ARNm. En réalité, ceci est dû à l'existence d'*interruptions qui séparent les différentes parties d'une région codante dans l'ADN*.

L'existence de gènes morcelés fut révélée grâce à des expériences permettant d'identifier l'ADN correspondant à un ARNm particulier. Le but d'une telle approche, aller de l'ARN au gène, était d'identifier des unités de transcription intactes comprenant les séquences flanquantes où figurent promoteurs et autres éléments régulateurs qui ne sont pas nécessairement présents dans l'ARNm.

Initialement, on supposait qu'un ARNm avait la même séquence que l'ADN à partir duquel il avait été transcrit. Afin de vérifier qu'une séquence d'ADN génomique correspondait effectivement à la séquence de l'ARNm qui avait permis son isolement, les deux séquences furent comparées soit par microscopie électronique, soit par cartographie de restriction. Cette expérience mit en évidence l'existence de régions supplémentaires dans l'ADN génomique.

La présence de séquences supplémentaires permet d'expliquer au moins une partie de la différence de taille entre ARNm et ARN nucléaire ; le transcrit primaire contient des séquences supplémentaires qui sont supprimées durant sa maturation en ARNm. Bien que faisant partie du gène de structure, ces séquences interposées n'ont pas de fonction codante.

Il convient maintenant de revoir la terminologie définie au Chapitre 4 afin de décrire la relation unissant un gène et son produit ARN correspondant. Un gène en mosaïque est constitué d'une alternance d'**exons** et d'**introns**. Les exons sont les séquences représentées dans l'ARN (qui peut être un ARNm, un ARNr ou un ARNt, ce qui signifie que les exons ne codent pas forcément pour une protéine). Les introns sont les séquences interposées qui sont supprimées lors de la maturation du transcrit primaire en ARNm. Un gène commence et finit toujours par un exon (ces exons correspondent aux extrémités 5' et 3' de l'ARN) ; il peut y avoir n'importe quel nombre d'introns entre eux.

Seuls les eucaryotes ont des séquences codantes interrompues, mais tous les gènes eucaryotes ne sont pas en mosaïque. La plupart des gènes chez les eucaryotes supérieurs sont en mosaïque mais il n'est pas possible de déterminer une relation entre la taille du gène et celle

de l'ARNm. Par conséquent, nous ne savons pas encore dans quelle mesure la présence d'introns permet de résoudre le paradoxe de la valeur C.

Une fois la structure du gène définie, le problème suivant est de déterminer son contexte. Quelles longueurs d'ADN de chaque côté du gène sont impliquées dans son fonctionnement ? A quelle distance se trouve le gène suivant, et quelle proportion du génome représentent les séquences qui sont situées entre les unités de transcription ? Est-ce que deux gènes adjacents ont tendance à interagir ? Existe-t-il un mécanisme qui permet de contrôler des régions chromosomiques de taille plus importante que de simples gènes ?

LES GÈNES EXISTENT SOUS TOUTES SORTES DE FORMES ET DE TAILLES

Quand un gène n'est pas en mosaïque, la carte de restriction de son ADN correspond exactement à celle de son ARNm (qui est obtenue grâce au transcrit inverse de l'ADNc). Quand un gène contient une séquence absente dans l'ARNm correspondant (ADNc), les cartes de restriction sont différentes.

Quand un gène possède un intron, la carte de chacune de ses extrémités correspond à celle des extrémités de la séquence du messager. A un certain point, à l'intérieur du gène, les cartes divergent et une région supplémentaire est présente dans le gène, contenant une série de sites de reconnaissance absente dans le messager. Sur la **Figure 18.1**, sont comparées les cartes de restriction du gène de la β -globine et de son ARNm. Deux introns sont mis en évidence. La cartographie de restriction permet de détecter des segments de gène supérieurs à 20-30 pb.

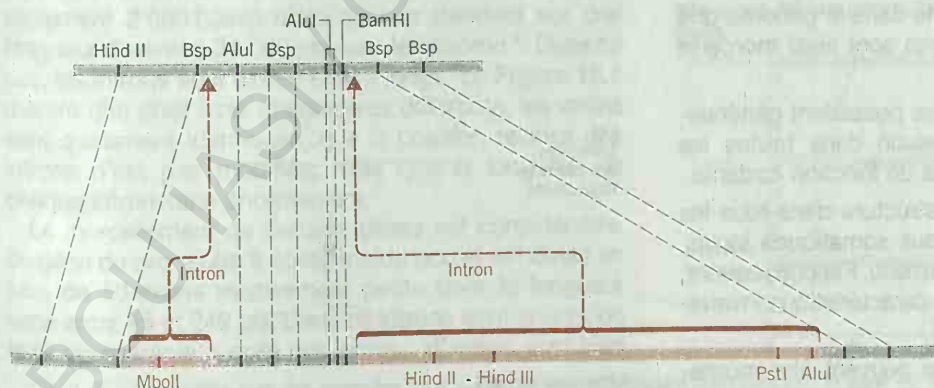


Figure 18.1
La comparaison des cartes de restriction de l'ADNc et de l'ADN génomique de la β -globine de souris montre que le gène possède deux régions supplémentaires, absentes dans l'ADNc. Les autres régions de l'ADNc et du gène sont strictement identiques.

Comme il est difficile de travailler directement avec l'ADN génomique, la caractérisation de gènes s'effectue généralement grâce à des séquences clonées. Une fois l'ADNc disponible, les techniques décrites dans le Chapitre 16 permettent d'obtenir les clones génomiques qui lui correspondent. Ces derniers peuvent ensuite être comparés avec l'ADNc par cartographie de restriction. La seule limite technique de cette approche est qu'il est difficile, voire impossible, d'isoler de longs gènes sous la forme d'un seul clone ; dans ce cas, les clones représentent chacun une partie du gène et doivent être caractérisés individuellement. Afin de s'assurer qu'il ne manque pas une partie du gène entre les séquences clonées, il est essentiel d'utiliser des clones dont les séquences génomiques se chevauchent.

En dernier lieu, la comparaison des séquences nucléotidiques des clones génomiques et de l'ADNc met en évidence la localisation et la longueur des introns. Cette analyse de la séquence est nécessaire pour s'assurer que tous les segments du gène ont été identifiés. En effet, dans plusieurs cas, le séquençage de l'ADN génomique a révélé la présence d'une courte séquence codante supplémentaire non identifiée par cartographie de restriction. Lorsqu'une séquence codante est inférieure à 50 pb, il peut lui être difficile de s'hybrider à une sonde ADNc, elle peut par conséquent passer inaperçue entre les séquences interposées situées de part et d'autre.

Aucune raison particulière n'a encore été décelée pour expliquer l'extrême variété de structure des gènes eucaryotes. La séquence des gènes non morcelés est colinéaire à celle de leur ARNm. La plupart des gènes eucaryotes sont morcelés, mais ils montrent une grande diversité autant dans le nombre que dans la taille de leurs introns.

D'importantes caractéristiques sont communes à tous les gènes morcelés :

- L'ordre dans lequel apparaissent les différentes parties d'un gène morcelé est le même dans le génome que dans l'ARNm mature. Les gènes sont ainsi morcelés plutôt que dispersés.
- Les introns des gènes nucléaires possèdent généralement des codons de terminaison dans toutes les phases de lecture, ils n'ont pas de fonction codante.
- Un gène morcelé a la même structure dans tous les tissus, lignée germinale et tissus somatiques inclus, qu'il soit exprimé ou non dans ce tissu. Par conséquent, la présence d'un intron est une caractéristique invariable.

Toutes les catégories de gènes peuvent être morcelées : les gènes nucléaires codant pour des protéines, les gènes nucléolaires codant pour les ARNr et les gènes

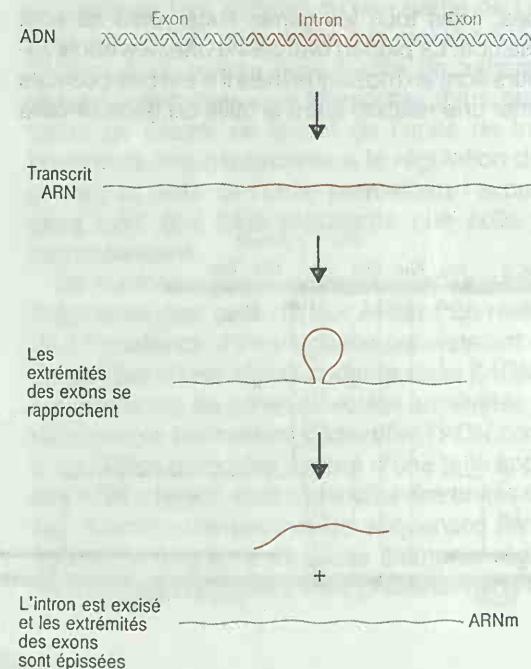
codant pour les ARNt. Des interruptions sont également trouvées dans les gènes mitochondriaux de levure et dans les gènes des chloroplastes. Il semble qu'aucune classe d'eucaryotes ne soit dépourvue de gènes morcelés, ces derniers ont même été trouvés chez archéobactérium et un phage d'*E. coli*. Ils semblent seulement être absents des génomes eubactériens.

Les introns doivent être supprimés avant qu'un gène ne soit traduit (ou avant qu'un ARNr ou un ARNt morcelé ne remplisse sa fonction). Ils sont présents dans le transcrit primaire mais absents de l'ARNm mature. Le processus de cette suppression est appelé **l'excision-épissage**. Il implique essentiellement l'excision précise d'un intron à partir du transcrit primaire ; après excision, les extrémités de l'ARN sont épissées afin que soit reformée une molécule intacte. Cette réaction est illustrée sur la **Figure 18.2** ; les mécanismes d'excision-épissage sont discutés dans le Chapitre 23.

Un gène de structure est défini comme étant la région dans le génome située entre les points correspondant aux bases 5' et 3' terminales de l'ARNm mature. Nous savons que la transcription commence à l'extrémité 5' de l'ARNm, mais il est possible qu'elle s'étende au-delà de l'extrémité 3' engendrée par clivage (voir Chapitre 24).

Figure 18.2

Un gène morcelé est constitué alternativement d'exons et d'introns ; les introns sont excisés du transcrit ARN et les exons sont épissés, ce qui engendre un ARNm constitué uniquement des séquences exoniques.



	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3
Longueur en pb	142-145	116-130	222	573-904	216-255
Correspondance	Région 5' non traduite + acides aminés 1 à 30		Acides aminés 31 à 104		Acides aminés 105 à la fin + région 3' non traduite

Figure 18.3

Tous les gènes fonctionnels de la globine ont une structure morcelée comportant trois exons. Les longueurs indiquées sur la Figure sont celles des gènes de la β -globine de mammifère.

La plupart des variations de longueur entre différents gènes de globine proviennent d'une variation de longueur dans le second intron. Chez la souris, le second intron du gène de l' α -globine est seulement long de 150 pb, la longueur totale du gène est par conséquent de 850 pb alors que celle du gène majeur de la β -globine est de 1382 pb. La variation de longueur des gènes est plus importante que la variation qui existe entre les longueurs d'ARNm (l'ARNm de l' α -globine est de 585 bases, celui de la β -globine de 620 bases).

La définition du gène peut être élargie si l'on y inclut les régions régulatrices : une unité de transcription est constituée d'un promoteur, d'autres régions de contrôle situées dans la région en amont du gène, du gène lui-même, et (parfois) d'une terminaison.

Certains gènes morcelés ne possèdent qu'un seul ou un petit nombre d'introns. Les gènes de la globine constituent un modèle qui a été beaucoup étudié (voir Chapitre 19). Les deux types de gènes de globine, alpha et bêta, ont le même type de structure. La cohérence de l'organisation des gènes de β -globine de mammifère est évidente si l'on regarde la **Figure 18.3**.

Les interruptions dans le gène apparaissent à des positions identiques (par rapport à la séquence codante) dans tous les gènes de globine actifs connus, c'est-à-dire ceux de plusieurs mammifères, d'oiseaux et de grenouille. Le premier intron est toujours relativement court, le second est généralement plus long, bien que les longueurs puissent varier.

Certains gènes sont bien plus importants que leur ARNm. Un cas extrême est illustré par la DHFR de mammifère (dihydrofolate réductase), dont le gène est constitué de six exons représentant un ARNm long de seulement 2 000 bases alors qu'ils s'étendent sur une longueur d'environ 31 000 pb dans le génome*. Dans ce cas, les introns sont extrêmement longs. La **Figure 18.4** montre que chez trois mammifères différents, les exons sont quasiment identiques, que la position relative des introns n'est pas modifiée, mais que la longueur de chaque intron varie énormément.

Le morcellement de certains gènes est considérable. Le gène du pro-alpha 2 collagène de poulet est divisé en plus de 50 exons relativement petits dont la longueur varie entre 45 et 249 pb. Certains introns sont courts, de taille comparable à celle des exons ; d'autres sont bien

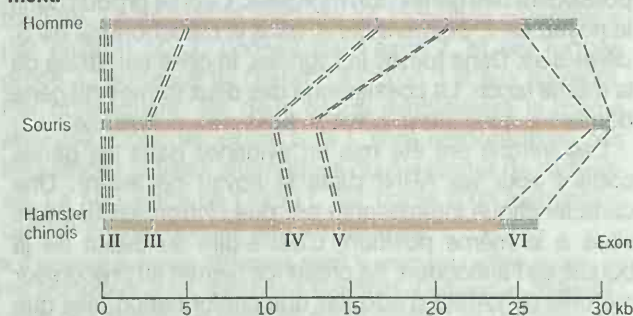
plus importants, ce qui a pour résultat de disperser les exons (qui représentent environ 5 000 pb au total) sur une longueur de 40 000 pb dans le génome.

Dans certains gènes de structure eucaryotes, les exons ne représentent qu'une petite proportion de la longueur totale du gène. Nous pouvons considérer un gène caractéristique comme étant constitué d'une série d'exons relativement courts distribués sur une grande région du génome.

Les introns représentent probablement une proportion plus importante du génome eucaryote que les exons. Chez les eucaryotes inférieurs, les introns sont probablement moins fréquents et les gènes ne sont sans doute pas morcelés, ou bien ne possèdent qu'un petit nombre d'introns. Il est certain que les exemples de complexité de structure les plus étonnants sont illustrés par les gènes d'eucaryotes supérieurs. La forme mosaïque du gène permet d'expliquer, dans une certaine mesure qui reste à définir, la faible proportion d'ADN représentée dans l'ARNm.

Figure 18.4

Les gènes de mammifères codant pour la DHFR ont tous la même organisation générale : ils sont composés de petits exons et de très longs introns dont la longueur peut cependant varier énormément.



* N. d. T. : Rappelons que des exemples plus extrêmes sont maintenant bien connus : le gène du facteur anti-hémophilique A est d'environ 180 kb et code pour un messager d'environ 9 kb ; le gène de thyroglobuline ferait plus de 300 kb pour un messager d'environ 8 kb ; enfin, le gène de dystrophine ferait environ 2 500 kb et donne un ARN mature de seulement 14 kb.

LES INTRONS DES GÈNES CODANT POUR LES ARNr ET LES ARNt

Dans certains cas, il existe des introns dans les gènes d'ARNr. En fait, les gènes morcelés furent découverts chez *D. melanogaster* où à peu près deux tiers des gènes de l'ARN 28S possèdent une interruption de 5 000 pb dans la séquence codante. (Les gènes codant pour les ARNr sont toujours présents en multiples copies le plus souvent identiques ; voir Chapitre 19.) Cependant les gènes morcelés ne sont pas utilisés pour la synthèse des ARNr ; les gènes actifs sont ceux qui ont une séquence codante continue.

A l'inverse, chez certains eucaryotes inférieurs, toutes les copies du gène codant pour le grand ARNr sont morcelées. Ce sont, par exemple, celles de *Tetrahymena pigmentosa* (un cilié) et de *Physarum polycephalum* (une moisissure de vase). Ces gènes sont, par conséquent, dans une situation analogue à celle des gènes de structure nucléaires non répétitifs : l'intron doit être excisé afin que le produit du gène soit actif. Toutefois, chez d'autres eucaryotes inférieurs, tels que *S. cerevisiae* (une levure) et *D. discoideum* (une moisissure de vase), aucun des gènes codant pour les ARNr n'est morcelé. Dans les mitochondries de *S. cerevisiae* et de *N. crassa* (un champignon) et dans les chloroplastes de *C. reinhardtii* (une algue), le gène codant pour le grand ARNr est morcelé, alors qu'il est intact dans les mitochondries et les chloroplastes d'autres espèces.

Il n'existe aucune relation systématique entre la distribution des gènes morcelés et non morcelés codant pour les ARNr. Bien qu'il n'y ait qu'un seul intron (sauf dans un cas) situé environ aux deux tiers du gène, le site précis où il interrompt la séquence codante diffère dans chaque cas. Il n'y a pas d'interruption dans la séquence codante des gènes codant pour les petits ARNr.

Dans deux cas, la présence de l'intron est polymorphique ; certaines lignées d'un organisme possèdent des gènes codant pour les ARNr morcelés alors que d'autres possèdent des gènes non morcelés. Ceci se produit dans le noyau de *T. pigmentosa* et dans la mitochondrie de *S. cerevisiae*. Dans toutes les lignées, le gène est utilisé de la même façon. La coexistence des deux formes du gène démontre que l'intron n'est ni essentiel, ni nuisible.

Des introns ont été mis en évidence dans les gènes codant pour les ARNt dans le noyau de levure. Une caractéristique intéressante est que l'intron est toujours situé à la même position, c'est-à-dire au début de la boucle de l'anticodon. Sa présence permet au précurseur morcelé de l'ARNt d'adopter une conformation telle que l'anticodon soit apparié avec une partie de l'intron. Cette étape fait probablement partie du mécanisme d'excision-épissage qui est illustré sur la Figure 23.1.

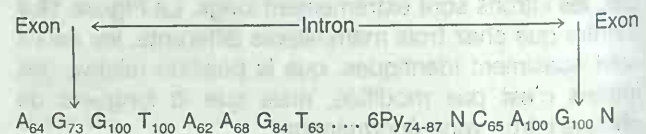
La longueur de l'intron varie de 14 à 46 pb dans les différents gènes codant pour les ARNt. La présence de l'intron ne peut être une caractéristique essentielle du gène de l'ARNt nucléaire de levure puisque certains gènes d'ARNt ne sont pas morcelés. Dans les familles de gènes d'ARNt, certains membres sont morcelés alors que d'autres sont continus.

LES JONCTIONS EXON-INTRON ONT UNE SÉQUENCE CONSENSUS

Il est possible de trouver les jonctions entre exons et introns en comparant les séquences nucléotidiques de l'ARNm et du gène de structure. Deux caractéristiques sont significatives. Premièrement, *il n'existe pas d'homologie ou de complémentarité entre les deux extrémités d'un intron*. L'hypothèse d'une liaison entre les deux extrémités de la structure secondaire, comme étape préliminaire de l'épissage, doit par conséquent être exclue. Deuxièmement, *les jonctions ont une séquence consensus bien conservée, bien que relativement courte*, qui s'avère nécessaire à l'épissage dans le noyau (voir Chapitre 23).

Bien souvent, la comparaison de l'ARNm au gène ne permet pas de déterminer un seul site de jonction exon-intron. La difficulté réside dans le fait qu'une courte séquence, généralement de 1 à 4 pb, est répétée à chaque extrémité de l'intron. Sur la **Figure 18.5** est illustré l'exemple de la β -globine de souris (ou de lapin), dans lequel chacune des 4 paires de sites pourrait définir les extrémités de l'intron.

Malgré cette difficulté, l'on peut désigner une extrémité spécifique à chaque intron en se basant sur l'homologie des jonctions exon-intron. Ces dernières doivent toutes être le plus proche possible de la séquence consensus ; dans ce schéma, les flèches désignent les extrémités supposées de l'intron :



(Dans ce cas comme dans beaucoup d'autres, nous n'écrivons que la séquence du brin d'ADN qui est identique à celle du produit ARN.)

Les indices indiquent, à chaque position consensus, le pourcentage d'apparition d'une base spécifique (ou d'un type de base). La région hautement conservée se situe dans l'intron au niveau des jonctions présumées. Ceci

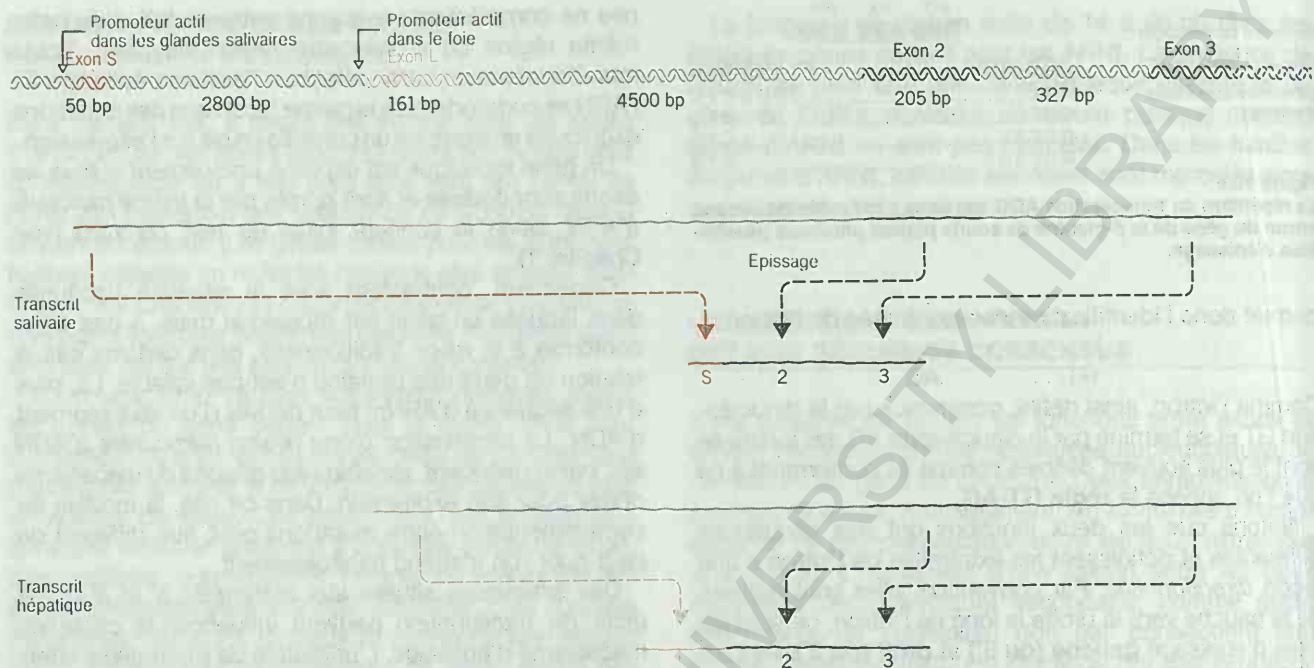


Figure 18.6

L'utilisation de promoteurs alternatifs, à partir du gène de l'amylase de souris, permet un épissage différentiel qui donne lieu à deux ARNm.

Un autre type de modification dans le mécanisme de maturation des ARNm est illustré par l'expression des gènes des immunoglobulines. Il consiste à substituer un exon par un autre à l'extrémité 3' de l'unité de transcription. Cela modifie la séquence des acides aminés C-terminaux de la protéine, ce qui a pour effet de remplacer la région responsable de l'attachement membranaire par une région permettant la sécrétion. Ce mécanisme est discuté dans le Chapitre 32.

Lors de l'expression tardive de l'adénovirus, la même séquence « leader » peut être épissée à plusieurs séquences codantes différentes. Sur la Figure 18.7, sont schématisées les différentes versions de ces événements.

L'expression d'une unité de transcription unique est initiée à un seul promoteur. La première partie de l'unité de transcription porte trois séquences qui sont épissées ensemble afin de former une séquence non transcrite (séquence « leader »). Les composants de ce « leader » sont relativement courts et sont indiqués comme étant les exons « leader » 1, 2 et 3.

En aval de ces séquences, plusieurs régions codantes sont présentes, chacune correspondant à une protéine

virale tardive. Elles sont indiquées comme étant les exons codants A, B, C, etc... Dans chaque transcrit primaire, le « leader » tripartite formé grâce aux deux premiers événements d'épissage est épissé à l'une des régions codantes.

L'extrémité 3' de chaque exon codant est engendrée par clivage du transcrit primaire et polyadénylation de l'extrémité terminale. Les séquences se situant plus en aval sont abandonnées. L'extrémité 5' de la région codante est ensuite épissée au « leader » tripartite. Pour le premier exon codant, cela ne pose aucun problème particulier, un intron est supprimé selon le mécanisme habituel. En ce qui concerne l'exon B, « l'intron » consiste en la région située entre son extrémité 5' et l'exon « leader » 3. Cette région comprend l'exon codant A. Par conséquent, une séquence donnée peut être considérée soit comme un intron, soit comme un exon selon le mécanisme de maturation choisi.

Seul l'un des exons alternatifs peut être utilisé dans une molécule d'ARN donnée. Les autres sont abandonnés s'ils sont situés en amont de l'exon utilisé. La production de l'extrémité 3' de l'ARN nucléaire détermine

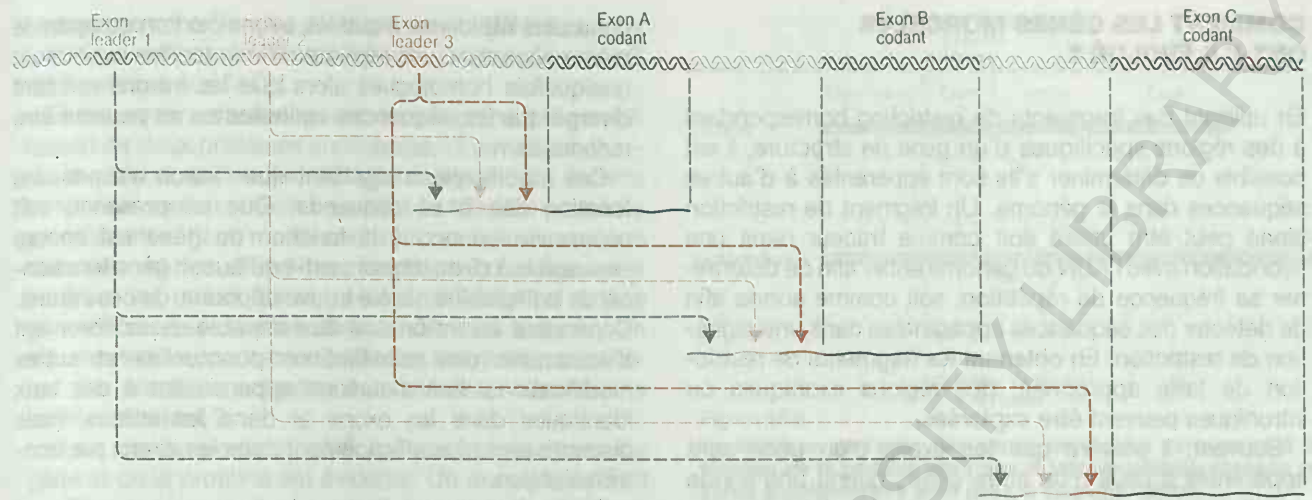


Figure 18.7

Le clivage alternatif donnant lieu à différentes extrémités 3' permet à une même séquence leader d'être épissée à différents messagers lors de l'expression tardive de l'adénovirus. Seuls trois exons sont illustrés dans la Figure, mais il existe en réalité cinq groupes de messagers définis par leur extrémité 3' commune, qui correspondent au minimum à neuf régions codantes différentes.

probablement l'exon allant être épissé avec le « leader » tripartite.

L'épissage alternatif (ou différentiel) peut donner lieu à deux protéines comportant des séquences communes et chevauchantes issues d'un même segment d'ADN. Le gène de la troponine T du muscle de rat en est un exemple. La **Figure 18.8** montre que la moitié 3' du gène comporte 5 exons, mais que seulement 4 sont utilisés pour la synthèse d'un ARNm. Trois exons, WXZ, sont identiques dans les deux modes d'expression. Cependant, dans un cas, l'exon alpha est épissé entre X et Z, dans l'autre c'est l'exon bêta qui est utilisé.

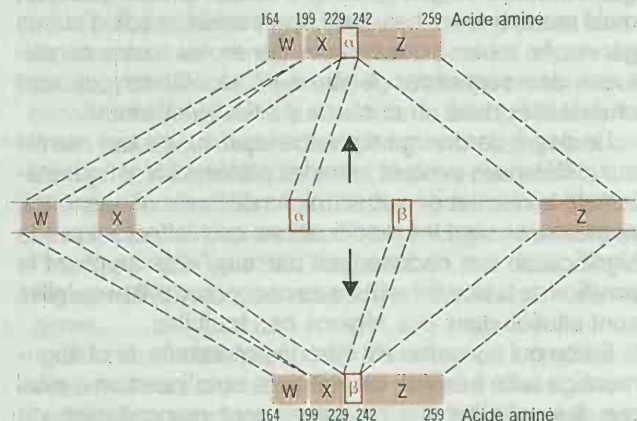
Par conséquent, les formes alpha et bêta de la troponine T diffèrent par la séquence des acides aminés comprise entre les codons 229 et 242, déterminée selon l'exon alternatif, alpha ou bêta, utilisé. L'un ou l'autre des exons alpha et bêta peut être utilisé pour la synthèse d'un ARNm, mais ils ne peuvent être utilisés ensemble pour la synthèse d'un même ARNm.

Les mécanismes d'épissage alternatif existent dans plusieurs gènes cellulaires ; la substitution d'un exon par un autre modifie la séquence des acides aminés et donne lieu à des protéines apparentées qui ont probablement des fonctions différentes. Cette utilisation économique de l'ADN est surprenante au regard de l'excès apparent d'ADN postulé par le paradoxe de la valeur C.

Dans un cas extrême chez la drosophile, la région codante d'un gène est entièrement présente sur un brin d'ADN, alors que sur l'autre brin, elle fait partie d'un intron dans un gène différent !

Figure 18.8

Les formes α et β de la troponine T proviennent de l'épissage différentiel d'un même gène.



COMMENT LES GÈNES MORCELÉS ONT-ILS ÉVOLUÉ ?

En utilisant des fragments de restriction correspondant à des régions spécifiques d'un gène de structure, il est possible de déterminer s'ils sont apparentés à d'autres séquences dans le génome. Un fragment de restriction cloné peut être utilisé soit comme traceur dans une hybridation avec l'ADN du génome entier afin de déterminer sa fréquence de répétition, soit comme sonde afin de détecter des séquences apparentées dans une digestion de restriction. En obtenant les fragments de restriction de taille appropriée, des régions exoniques ou introniques peuvent être explorées.

Souvent, il s'avère que les exons d'un gène sont apparentés à ceux d'un autre gène ; quand une sonde exonique est utilisée, elle permet la détection de fragments qui font partie d'un ou de plusieurs autres gènes. Cela signifie que les deux gènes proviennent par duplication d'un gène ancestral commun, dont les copies ont par la suite accumulé des différences l'une par rapport à l'autre.

Généralement, les introns ne sont pas apparentés à d'autres séquences. Quand deux gènes sont des exons apparentés, le degré d'homologie est moins important dans les introns que dans les exons. Étant donné que les exons doivent coder pour des séquences d'acides aminés, leur potentiel de changement de séquence est limité. Libres de cette contrainte, les introns peuvent subir et accumuler plus de substitutions que les exons.

La définition initiale de l'ADN non répétitif le considère comme étant constitué de séquences uniques dans le génome. Les introns sont plus proches de cette définition que les exons. La réponse à la question « les gènes de structure sont-ils non répétitifs ? » est ambiguë : la longueur totale d'un gène peut être unique en tant que telle, mais ses exons sont souvent apparentés à ceux d'autres gènes. Au moins pour certains gènes, les exons constituent des séquences légèrement répétitives qui sont enchâssées dans un contexte particulier d'introns.

Le degré de divergence entre deux exons est corrélé aux différences existant entre les protéines. Il est généralement le résultat de substitutions de bases. Cependant, nombreuses sont les modifications qui n'affectent pas la signification des codons, soit parce qu'elles touchent la position de la troisième base du codon, soit parce qu'elles sont situées dans des régions non traduites.

En ce qui concerne les introns apparentés, le changement de taille (résultat de délétions ou d'insertions) ainsi que les substitutions de bases sont responsables du mécanisme de divergence. Les introns évoluent beau-

coup plus rapidement que les exons ; si l'on compare le même gène dans des espèces différentes, les exons sont quelquefois homologues alors que les introns ont tant divergé que les séquences équivalentes ne peuvent être reconnues.

Ces modifications signifient que l'intron n'a pas une fonction liée à sa séquence. Que sa présence soit nécessaire ou non à la fonction du gène est encore impossible à dire ; toujours est-il qu'aucun gène fonctionnel de la β -globine n'a été trouvé dépourvu de ces introns. Cependant un intron doit être capable assez librement d'accumuler des substitutions ponctuelles et autres modifications. Les mutations apparaissent à des taux identiques dans les exons et dans les introns, mais disparaissent plus efficacement dans les exons par contre-sélection.

Quelle était la structure originelle des gènes qui sont aujourd'hui morcelés ? Les unités ancestrales codant pour les protéines consistaient-elles en des séquences d'ADN non morcelées, dans lesquelles des introns furent ultérieurement insérés au cours de leur évolution vers la forme actuelle ? À l'inverse, ces gènes n'avaient-ils pas à l'origine une structure morcelée qui a été maintenue jusqu'à aujourd'hui ? En d'autres termes, l'on peut se demander si la différence entre gènes procaryotes et eucaryotes est le résultat de l'acquisition d'introns chez les eucaryotes ou de la perte d'introns chez les procaryotes.

La structure mosaïque pourrait-elle être le reliquat d'un ancien mécanisme de construction de gènes ayant pour but de produire de nouvelles protéines ? Supposons qu'une cellule primitive ait un nombre précis de séquences différentes codant pour des protéines. Un aspect de son évolution serait probablement de reconstituer et de juxtaposer ces différentes unités polypeptidiques dans le but de construire de nouvelles protéines.

Si une unité codant pour une protéine doit représenter une série continue de codons, chacune de ces reconstitutions doit être l'objet d'une recombinaison précise de l'ADN, ceci afin que deux unités soient bien placées l'une par rapport à l'autre, extrémité contre extrémité dans la même phase de lecture. En outre, si l'association n'est pas réussie, la cellule en reste endommagée car les deux unités initiales ne sont plus fonctionnelles.

Au contraire, si une recombinaison approximative de l'ADN place les deux unités codantes à l'intérieur d'une même unité de transcription, les mécanismes d'épissage peuvent alors intervenir et tenter de combiner les deux protéines dans le but d'obtenir une chaîne polypeptidique unique. Si ces combinaisons ne sont pas réussies, les unités codantes initiales restent disponibles pour d'autres tentatives. Une telle approche permet à la cellule d'es-

sayer et de contrôler des délétions dans l'ARN sans avoir à subir les effets de l'instabilité d'un tel mécanisme s'il était pratiqué sur l'ADN.

Si les protéines actuelles ont évolué grâce à la combinaison de deux protéines ancestrales, la formation d'unités a sans doute eu lieu de façon séquentielle dans le temps, un seul exon étant ajouté à la fois aux protéines préexistantes. Si ce modèle est vrai, nous pouvons nous demander si les différentes fonctions pour lesquelles codaient les unités qui ont permis la formation de ces gènes sont discernables dans la structure actuelle. En d'autres termes, peut-on identifier les fonctions particulières des protéines actuelles en s'intéressant aux exons individuellement ?

Dans certains cas, la relation entre les structures du gène et de la protéine est évidente. Un exemple édifiant est illustré par les immunoglobulines qui sont codées par des gènes dont chaque exon correspond exactement à un domaine fonctionnel connu de la protéine (voir Chapitre 32).

Dans d'autres exemples, certains exons peuvent être associés à des fonctions particulières. Dans les protéines sécrétées, le premier exon, codant pour la région N-terminale du polypeptide, détermine la séquence signal impliquée dans la sécrétion membranaire. L'insuline en est un exemple.

P parfois, l'évolution d'un gène passe par la duplication d'exons, créant ainsi une séquence répétitive dans la protéine. Dans le collagène de poulet, un exon de 54 pb a été multiplié de nombreuses fois, donnant lieu à une série d'exons d'une longueur variant de 54 pb à un multiple de 54 pb.

Un cas fascinant illustrant la conservation au cours de l'évolution est représenté par les globines, dont tous les gènes sont constitués de trois exons. Les deux introns sont localisés à des positions constantes par rapport à la séquence codante. L'exon central correspond au domaine de liaison à l'hème de la chaîne de globine. La protéine active est un tétramère constitué de deux chaînes de globine de type alpha et de deux autres de type bêta.

Deux autres types de protéine sont apparentés à la globine et nous permettent d'avoir une vision différente de sa structure. La myoglobine est une protéine animale monomérique de liaison à l'oxygène ; la comparaison de sa séquence d'acides aminés avec celles des sous-unités de la globine suggère qu'elles ont une origine commune quoiqu'ancienne. Les leghémoglobines sont des protéines de liaison à l'oxygène présentes dans la classe des légumineuses chez les plantes ; comme la myoglobine, elles sont monomériques. Elles ont toutes deux une origine commune avec les autres protéines de liaison à

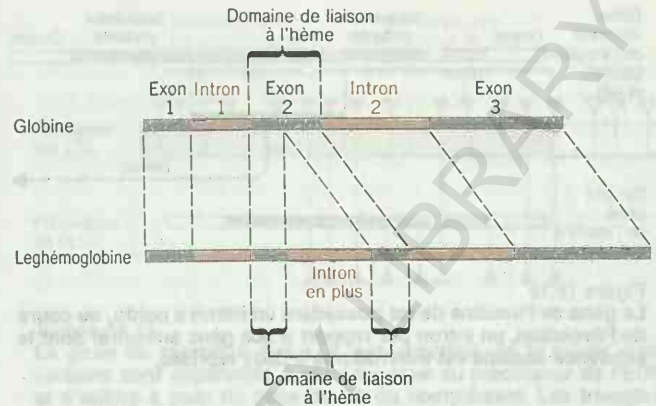


Figure 18.9

La structure de l'exon des gènes de globine correspond à la fonction de la protéine alors que la leghémoglobine possède un intron supplémentaire situé dans le domaine central.

l'hème. Les globines, myoglobine et leghémoglobine constituent la « super-famille » des globines, c'est-à-dire un ensemble de familles de gènes descendant tous d'un ancêtre commun.

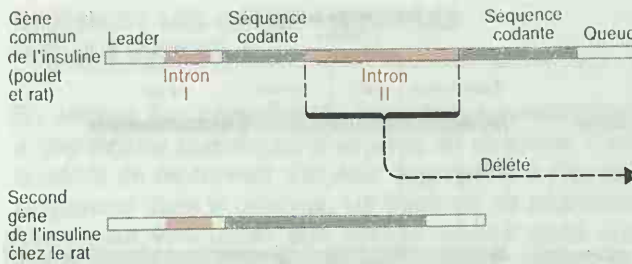
La myoglobine est codée dans le génome humain par un gène unique dont la structure est identique à celle des gènes de globine. Par conséquent, la structure en trois exons existait avant qu'évoluent séparément les globines et myoglobine.

Les gènes de leghémoglobines sont constitués de trois introns, le premier et le dernier sont situés sur la séquence codante aux mêmes niveaux que les introns des gènes de globine. A cause de cette remarquable similitude, il semble que les protéines de liaison à l'hème aient une origine commune extrêmement ancienne sous la forme d'un gène divisé, comme l'illustre la Figure 18.9.

L'exon central de la leghémoglobine sépare deux exons qui codent ensemble pour la séquence correspondant à l'exon central de la globine. L'exon central du gène de la globine dérive-t-il d'un gène ancestral par la fusion de ses deux exons centraux, liant ainsi les séquences codant pour les deux parties de la chaîne protéique qui constituent à elles deux la structure de liaison à l'hème ?

Les cas dans lesquels des gènes homologues ont des structures différentes devraient donner des informations quant à leur évolution. L'insuline en est un exemple. Les mammifères et les oiseaux ont un seul gène codant pour l'insuline, alors que les rongeurs en ont deux. Sur la Figure 18.10, sont représentées les structures de ces gènes.

Chez le poulet, le gène de l'insuline est constitué de deux introns ; chez le rat, seul l'un des deux gènes a cette même structure. Cette similarité signifie que le gène ancestral de l'insuline possédait deux introns. Cependant,

**Figure 18.10**

Le gène de l'insuline de rat possédant un intron a perdu, au cours de l'évolution, un intron par rapport à son gène ancestral dont la séquence codante est interrompue à deux reprises.

le second gène chez le rat n'est constitué que d'un intron. Pour expliquer sa structure, il faut imaginer le processus suivant : d'une part la duplication du gène ancestral chez les rongeurs, et d'autre part l'excision précise d'un intron dans l'une des deux copies.

L'organisation de certains gènes est très différente entre les espèces. Dans ces cas là, d'importantes excisions ou insertions d'introns doivent avoir eu lieu au cours de l'évolution.

L'exemple le mieux caractérisé est illustré par les gènes d'actine. Le gène d'actine typique a un « leader » non traduit d'une longueur supérieure à 100 bases, une région codante d'environ 1 200 bases et une queue d'environ

200 bases. La plupart des gènes d'actine sont morcelés ; les positions des introns peuvent être déterminées par rapport à la séquence codante (sauf pour un intron parfois trouvé dans le « leader »).

La **Figure 18.11** montre que pratiquement chaque gène d'actine est différent des autres en ce qui concerne l'organisation des introns. Si l'on prend en compte tous les gènes, les introns sont situés à 12 positions différentes. Cependant, aucun des gènes ne comporte plus de 6 introns ; certains gènes ne possèdent qu'un intron, et l'un d'entre eux n'est pas morcelé. L'on trouve des relations entre différentes espèces ; par exemple, 4 des positions introniques sont identiques chez le rat, le poulet et l'oursin de mer.

Comment en est-on arrivé à cette situation ? Si l'on suppose que le gène primitif d'actine était morcelé, et que tous les gènes d'actine actuels lui sont apparentés mais ont perdu des introns, chaque branche de l'évolution aurait perdu des introns différents. Certains introns ont probablement été entièrement perdus, le gène primitif devait donc en posséder au moins 20.

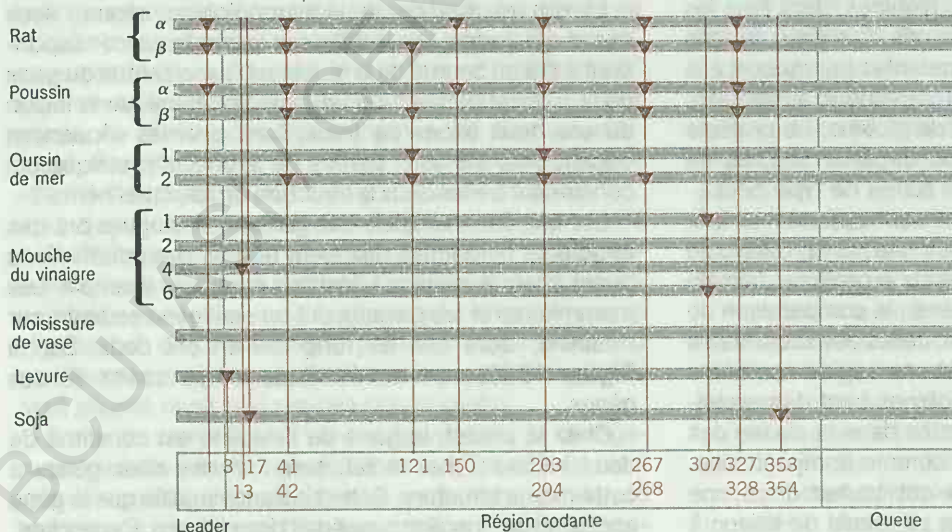
Le fibrinogène est dans une situation similaire, voire plus complexe, étant donné que trois gènes, liés dans le génome de rat, ont chacun un diagramme d'interruptions différent. Sur un total de 14 introns pour les trois gènes, 11 apparaissent à des sites uniques dans l'un des trois gènes.

Le génome mitochondrial représente un autre exemple

Figure 18.11

L'organisation des gènes d'actine varie énormément entre les espèces.

Les sites où se situent les introns sont indiqués par des triangles rouges ; chaque chiffre correspond au numéro du codon que chaque intron interrompt.



étonnant. Les gènes de levure et de mitochondries de mammifères codent pour les mêmes protéines mitochondriales, malgré une grande différence dans leur organisation. Les génomes mitochondriaux de vertébrés sont très petits, avec une organisation extrêmement compacte de gènes continus (voir Chapitre 20), alors que les génomes mitochondriaux de levure sont plus grands et constitués de gènes morcelés beaucoup plus complexes. Quelle est la forme ancestrale ?

La présence d'introns dans la mitochondrie de levure est-elle le reflet du fait que les mitochondries sont le résultat d'un événement ancien endosymbiotique qui a permis l'insertion d'un prototype bactérien dans le cytoplasme eucaryote ? Cela ne pose pas de problème si l'on suppose que l'insertion est antérieure à la perte des introns dans la lignée procaryote, bien qu'il faille alors également supposer que la capacité d'exciser des introns ait été perdue indépendamment dans les bactéries et les mitochondries de vertébrés, au cours de l'évolution.

Le polymorphisme semble être chose courante pour les gènes des ARNr et ARNt, qui sont alternativement trouvés sous des formes avec ou sans introns. Dans le cas des ARNt pour qui toutes les molécules sont conformes à une même structure générale, il semble improbable que l'évolution ait mis côte à côte deux régions du gène car enfin, les différentes régions sont impliquées dans l'appariement de bases qui donne la signification de la structure. Dans ce cas, les introns pourraient avoir été insérés dans des gènes continus.

Les séquences qui ont en commun des gènes apparentés en partie correspondent probablement à des exons qui ont migré ou ont été recrutés par d'autres gènes. La **Figure 18.12** résume les relations qui existent entre le récepteur au LDL humain (lipoprotéine plasmatique de faible densité) et d'autres protéines.

Au milieu du gène du récepteur au LDL se trouve une série d'exons apparentés aux exons du gène du précurseur de l'EGF (facteur de croissance de l'épiderme). Cinq des neuf introns de cette région sont à des positions identiques, confirmant l'origine commune des deux séries d'exons.

Dans la partie N-terminale de la protéine, une série d'exons code pour une séquence apparentée à une protéine du sang, le facteur C9 du complément. Nous nous attendons à ce que ces exons soient apparentés aux exons du C9. Le gène du récepteur au LDL est donc peut-être le résultat de l'assemblage de modules, étant également utilisés dans d'autres protéines.

La relation qui existe entre les exons et les domaines protéiques est quelque peu erratique en ce qui concerne les gènes connus. Dans certains cas, il y a une évidente

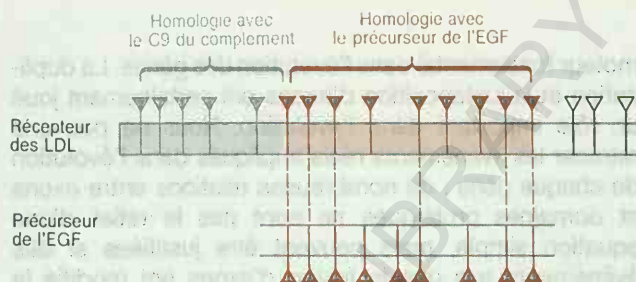


Figure 18.12

Le gène du récepteur des LDL est constitué de 18 exons dont certains sont apparentés à ceux du gène du précurseur de l'EGF et d'autres à ceux du gène du C9 du complément. Les triangles correspondent à la position des introns. Seuls certains introns de la région apparentée au récepteur de l'EGF ont une position identique à celle des introns du récepteur de l'EGF.

relation 1 : 1 ; dans d'autres, aucun modèle n'est discernable. Il se pourrait que l'excision d'introns ait permis la fusion d'exons adjacents. Une des difficultés de cette hypothèse est qu'il faut supposer que l'excision des introns se produit de façon précise, et n'altère pas l'intégrité de la région codante. L'on peut également imaginer que les introns se sont insérés dans des domaines codant pour un domaine protéique donné ; il faudrait alors supposer que l'intron a amené avec lui la capacité d'être excisé.

Les exons tendent à être de petite taille, codant en moyenne pour 20 à 80 acides aminés. Cela correspond à peu près à la taille du plus petit polypeptide pouvant avoir une structure repliée stable, c'est-à-dire à 20 à 40 résidus. Les protéines ont peut-être été initialement assemblées à partir de « modules » relativement petits. Chaque module ne correspondrait pas nécessairement à une fonction actuelle ; plusieurs modules pourraient se combiner pour assurer une fonction. Le nombre d'exons dans un gène tend à augmenter avec la longueur de sa protéine, ce qui est en accord avec le fait que les protéines pourraient avoir acquis des fonctions multiples par ajouts successifs des modules adéquats.

Cela peut éventuellement expliquer une autre caractéristique de la structure protéique : il semble que les sites correspondant aux liaisons exon-intron soient souvent situés à la surface de la protéine. Au fur et à mesure que des modules sont ajoutés à la protéine, les liaisons exon-intron des modules récemment ajoutés tendent à être à la surface.

La relation qui existe entre certains exons et certains domaines protéiques corrobore l'idée qu'elle fut un

moteur fondamental dans l'évolution des gènes. La duplication et la juxtaposition d'exons ont certainement joué un rôle important dans l'évolution. Nous ne pouvons retracer les événements réels impliqués dans l'évolution de chaque gène ; de nombreuses relations entre exons et domaines protéiques ne sont pas le reflet d'une équation simple, mais peuvent être justifiées si des événements tels que la fusion d'exons ont modifié la structure ancestrale au cours de l'évolution des gènes nucléaires.

POUR EN SAVOIR PLUS

Une discussion plus approfondie sur la structure des gènes morcelés fut réalisée dans *Gene Expression de Lewin, 2, Eucaryotic Chromosomes* (Wiley, New York, 1980, pp. 790-847). D'autres éléments furent apportés dans *Cell* (22, 645-646, 1980). La première revue générale sur la structure des gènes depuis la découverte des gènes morcelés a été réalisée par **Breathnach & Chambon** (*Ann. Rev. Biochem.* 50, 349-383, 1981). La relation unissant l'organisation des gènes et la structure des protéines a été analysée par **Blake** (*Int. Rev. Cytol.* 95, 149-185, 1985).

CHAPITRE 19

LES GÈNES DE STRUCTURE APPARTIENNENT À DES CLASSES DE TAILLE DIFFÉRENTES

PARTIE 6

DES GROUPES DE SÉQUENCES APPARENTÉES

Ce que l'on trouve inscrit dans le programme génétique est par conséquent le résultat de toutes les reproductions passées, puisque toutes les traces des échecs ont disparu. Le message génétique, programme de l'organisme actuel, ressemble ainsi à un texte sans auteur qu'un relecteur aurait corrigé depuis plus de deux milliards d'années, l'améliorant en permanence, le raffinant et le complétant, éliminant progressivement toutes les imperfections.

François Jacob, 1973

CHAPITRE 19

LES GÈNES DE STRUCTURE APPARTIENNENT À DES FAMILLES DE TAILLES DIFFÉRENTES

Peu de gènes sont uniques dans le génome des eucaryotes. Bien que la plupart des gènes apparaissent dans l'ADN non répétitif, en isolant des fragments génomiques, nous trouvons souvent d'autres séquences ressemblant à leurs exons. Ces séquences sont les exons d'autres gènes, représentant des protéines identiques ou apparentées à celles codées par le premier. Les séquences des introns des gènes apparentés de la sorte diffèrent généralement beaucoup plus que celles des exons correspondants.

Considérant la structure interrompue, parfois largement dispersée, des gènes eucaryotes, nous pouvons décrire le génome eucaryote comme un océan d'introns (dont la séquence est principalement mais pas exclusivement unique) et sur lequel des îlots d'exons (parfois très courts) sont alignés en archipels distincts qui constituent des gènes.

L'existence répandue, de gènes dont les exons correspondants sont apparentés, renforce l'importance de la duplication en tant que mécanisme servant à produire de nouveaux gènes. Une copie peut évoluer via une mutation tandis que l'autre garde sa fonction d'origine.

L'histoire d'un gène comporte probablement une série d'événements parmi lesquels tout d'abord les exons le composant ont été rassemblés pour former une région codante ; ensuite la série complète d'exons et d'introns constituant le gène pourrait avoir été dupliquée. La duplication serait moins divergente au niveau de la

séquence des exons et plus divergente au niveau de celle des introns. A partir des séquences de gènes apparentés il est possible d'essayer de reconstituer l'histoire de leur évolution.

Une série de gènes issue par duplication et mutation d'un gène ancestral est appelée **famille multigénique**. Ses membres peuvent être groupés ou dispersés sur différents chromosomes (ou les deux à la fois).

Dans certains cas, les membres d'une famille multigénique sont demeurés identiques. Le regroupement est une condition au maintien de l'identité entre les gènes alors que des gènes groupés ne sont pas nécessairement identiques. La taille des **batteries de gènes (ou groupes)** peut varier entre les situations extrêmes où, d'une part, une duplication a produit deux gènes apparentés adjacents et où, d'autre part, des centaines de gènes identiques sont alignés en tandem.

Le fait que des gènes voisins soient dispersés dans différents emplacements peut provenir de la **translocation** d'un gène, postérieurement à une duplication. Après leur séparation, les séquences des gènes, habituellement, divergent.

Les membres d'une famille de gènes de structure ont des fonctions voisines ou identiques, bien qu'ils puissent s'exprimer à des moments différents ou dans différents types de cellules.

Une importante répétition en tandem d'un gène peut survenir lorsque le produit est nécessaire en quantités

exceptionnellement abondantes. Les gènes codant pour l'ARN ribosomal (ARNr) ou les histones en sont l'exemple. Un degré moindre de répétition peut être utilisé pour produire des protéines légèrement différentes dans des circonstances particulières. Par exemple, différentes globines sont produites à l'usage des globules rouges adultes et embryonnaires, alors que, par ailleurs, différentes actines sont utilisées dans des cellules musculaires et non musculaires.

Quelquefois, la répétition ne laisse entrevoir aucune signification ; par exemple, nous n'avons connaissance d'aucune différence dans l'expression ou la fonction des gènes dupliqués de l'insuline des rongeurs (un gène unique pour l'insuline est en effet suffisant chez les autres mammifères).

Parfois, tous les membres de la famille sont des gènes fonctionnels, plus ou moins éloignés dans le degré d'évolution ; d'autres fois, certains sont des **pseudogènes** non fonctionnels, reliques de l'évolution, issus de gènes qui ont dû être fonctionnels, mais qui sont devenus inactifs et ont accumulé de nombreuses mutations.

LES GÈNES DE LA GLOBINE SONT ORGANISÉS EN DEUX GROUPES

Le transport de l'oxygène dans le flux sanguin est une fonction essentielle dans le royaume animal, codée par une ancienne famille génique. Le constituant principal du globule rouge est le tétramère de globine, associé à son groupement hème (liant le fer) sous forme d'hémoglobine. Les gènes fonctionnels de la globine, dans toutes les espèces, ont la même structure générale, constituée de trois exons comme nous l'avons décrit précédemment (Figure 18.9). Nous avons conclu que tous les gènes de la globine dérivait d'un seul gène ancestral ; ainsi, en comparant le développement individuel des gènes de la globine à l'intérieur et entre les espèces, nous pouvons apprendre beaucoup sur les mécanismes impliqués dans l'évolution des systèmes géniques.

Dans les globules adultes, le tétramère de globine est constitué de deux chaînes α identiques et de deux chaînes β identiques. Les gènes de la globine α et β sont codés par des loci indépendants dont l'expression doit être coordonnée pour obtenir une production égale des deux polypeptides. Ce système fournit donc un exemple de la nécessité d'un contrôle simultané de gènes dispersés pour produire un phénotyp-e cellulaire spécifique.

Les cellules sanguines embryonnaires contiennent des tétramères d'hémoglobine différents de ceux de la forme

Tableau 19.1

Modifications des hémoglobines humaines pendant le développement.

Étape du développement	Hémoglobines
Embryonnaire (jusqu'à 8 semaines)	$\zeta_2\epsilon_2$ et $\zeta_2\gamma_2$ et $\alpha_2\epsilon_2$
Fœtal	$\alpha_2\gamma_2$
Adulte	$\alpha_2\delta_2$ et $\alpha_2\beta_2$

ζ est la première chaîne de type α exprimée, mais elle est rapidement relayée par α elle-même. Dans la voie β , ϵ et γ sont exprimées en premier lieu, δ et β les remplaçant ensuite. Chez les adultes, la forme $\alpha_2\beta_2$ représente 97 % de l'hémoglobine, $\alpha_2\delta_2$ approximativement 2 %, et environ 1 % est dû à la persistance de la forme fœtale $\alpha_2\gamma_2$.

adulte. Chaque tétramère contient deux chaînes de type α identiques et deux chaînes de type β , chacune d'elles est apparentée au polypeptide adulte et est plus tard remplacée par lui. Ceci est un exemple de contrôle du développement dans lequel différents gènes sont successivement exprimés ou réprimés pour fournir différents produits accomplissant la même fonction à différents moments.

Les détails de la parenté existant entre les hémoglobines embryonnaires et adultes varient selon l'organisme. La voie humaine consiste en trois étapes (qui ne sont pas encore bien délimitées). Zêta et alpha sont les deux chaînes de type α . Epsilon, gamma, delta et bêta sont les chaînes de type β . Les chaînes exprimées aux différentes étapes du développement sont résumées dans le **Tableau 19.1**.

La division des chaînes de globine en chaînes de type α et β reflète l'organisation des gènes. Chaque type de globine est codé par des gènes organisés en un groupe unique. Chez l'homme le groupe α est localisé sur le chromosome 16, et le groupe β sur le chromosome 11. Les structures des deux groupes dans le génome humain sont représentées dans la **Figure 19.1**.

S'étendant sur plus de 50 kb, le groupe β comporte cinq gènes fonctionnels — ϵ , deux γ , δ et β — et un pseudogène. Les deux gènes γ diffèrent dans leur séquence codante par seulement un acide aminé ; le variant G possède une glycine en position 136, là où le variant A possède une alanine.

Le groupe α , plus compact, est réparti sur plus de 28 kb et comporte un gène actif ζ , un pseudogène ζ , deux gènes α et deux pseudogènes α^* . Les deux gènes α codent pour la même protéine. Deux gènes identiques (ou plus) présents sur le même chromosome sont décrits comme des copies **non alléliques**.

Une organisation générale similaire se retrouve dans

* N. d. T. : Il existe, à l'extrémité 3' du groupe α , un gène supplémentaire appelé θ qui semble faiblement exprimé dans les cellules érythroïdes. La chronologie de son activation au cours du développement reste floue.

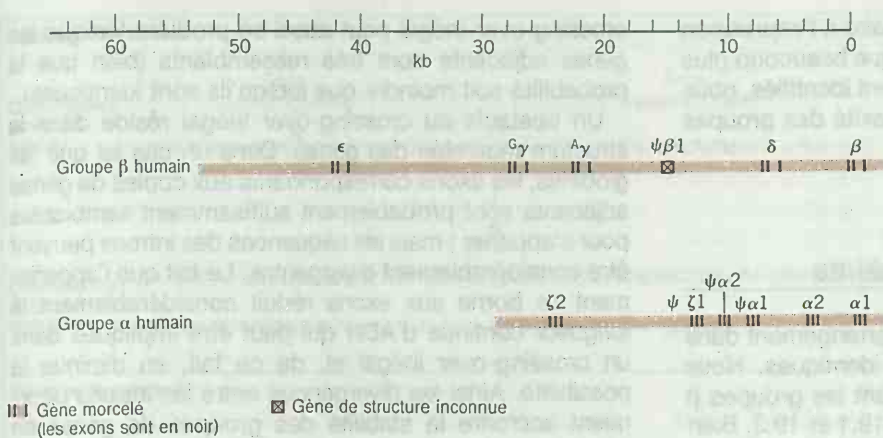


Figure 19.1

Chaque famille de gènes de globine humaine de type α et β est organisée en un groupe unique qui comporte des gènes fonctionnels et des pseudogènes (ψ).

L'organisation de ces gènes chez les autres primates supérieurs est très semblable. Par exemple, les groupes β du gorille ou du babouin sont pratiquement indiscernables de celui de l'homme. Tous les gènes sont transcrits de gauche à droite.

d'autres familles de gènes de globine de vertébrés, mais des détails de types, quantités et organisations des gènes peuvent varier. Quelques exemples de groupes de gènes β globine sont illustrés dans la Figure 19.2.

La caractérisation de ces groupements de gènes constitue un point général important. Il semble y avoir plus de membres d'une famille de gènes, à la fois fonctionnels et non fonctionnels, que nous ne l'aurions supposé d'après une analyse des protéines. Les gènes fonctionnels supplémentaires doivent représenter des duplications qui codent pour des polypeptides identiques ; ou bien ils peuvent être apparentés à des protéines connues, bien que différentes d'elles, et être vraisemblablement exprimés très brièvement ou en petites quantités. Les pseudogènes montrent des parentés variables avec les gènes fonctionnels, et ne peuvent être analysés qu'au niveau de la séquence de l'ADN. Si l'on prend ces faits en considération, il est difficile d'être certain que tous les membres d'un groupe ont été identifiés tant que les

régions avoisinantes n'ont pas été analysées au-delà des membres terminaux.

Tous les gènes de globine sont-ils englobés dans des groupes ? Dans chaque génome, tous les gènes actifs de globine se trouvent dans les groupes α et β, associés parfois à des pseudogènes. Dans le génome de souris, quelques pseudogènes supplémentaires ont été trouvés en dehors des groupes, localisés sur différents chromosomes. Ils doivent provenir de translocations qui ont fait sortir le gène du groupe.

En ce qui concerne la quantité d'ADN nécessaire pour coder pour une fonction particulière, nous savons que le fait de coder pour les globines de type β exige une longueur de 14 à 50 kb chez différents mammifères et chez un oiseau. Ceci est beaucoup plus important que l'on pourrait s'y attendre en examinant les β globines connues. Il est peu probable que d'autres protéines soient codées à l'intérieur du groupe β, mais nous ignorons encore la proportion d'ADN non codant (adja-

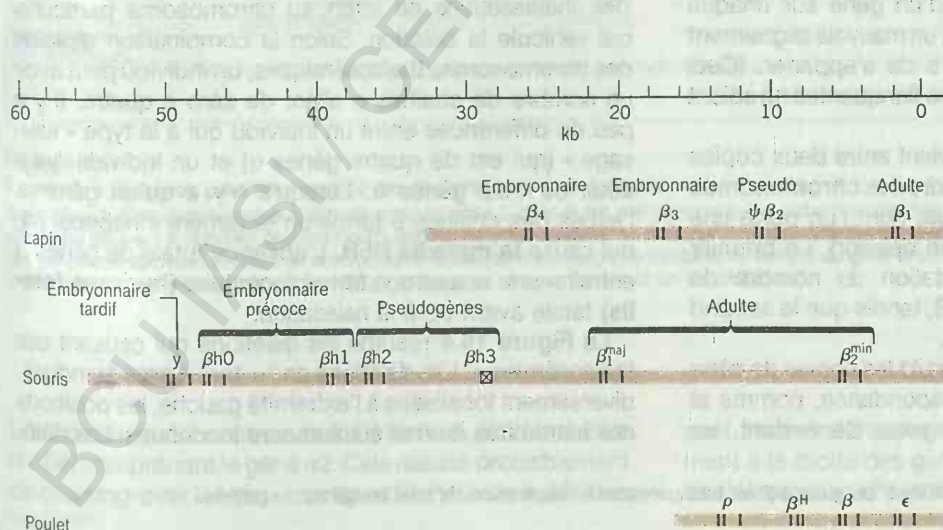


Figure 19.2

Groupes des gènes et pseudogènes de β globine présents chez d'autres vertébrés.

Le lapin possède quatre gènes de type β : deux embryonnaires, un pseudogène et un adulte répartis dans l'ordre de leur expression. Sept gènes de type β ont été dénombrés chez la chèvre. Parmi les sept gènes de la souris, on trouve deux gènes embryonnaires précoces, deux embryonnaires tardifs, deux adultes et deux pseudogènes.

Les groupes peuvent être plus courts et ne contiennent pas nécessairement des pseudogènes. Chez le poulet, le groupe β globine mesure moins de 14 Kb et semble ne contenir que quatre gènes fonctionnels de type β globine. Les deux gènes externes sont embryonnaires et les deux gènes internes sont adultes.

cent et intronique) fonctionnel, nécessaire à l'expression du gène de la β globine *. Jusqu'à ce que beaucoup plus de gènes ou de groupes de gènes soient identifiés, nous serons incapables de dire si la complexité des groupes de globine est caractéristique.

UN CROSSING-OVER INÉGAL RÉARRANGE LES GROUPES DE GÈNES

Il existe de fréquentes occasions de réarrangement dans un groupe de gènes apparentés ou identiques. Nous pouvons voir les résultats en comparant les groupes β de mammifères cités dans les Figures 19.1 et 19.2. Bien qu'ils aient tous la même fonction et la même organisation générale, chacun diffère par sa taille ; il existe une variation du nombre total et des types de gènes de β globine ; le nombre et les structures des pseudogènes sont différents. Tous ces changements ont dû apparaître depuis la divergence des mammifères, il y a 85 millions d'années (la dernière période d'évolution commune à tous les mammifères).

La comparaison montre qu'en règle générale, la duplication, le réarrangement et la variation d'un gène peuvent être un facteur d'évolution aussi important que la lente accumulation de mutations ponctuelles dans des gènes individuels. Quels sont les types de mécanismes responsables de la réorganisation des gènes ?

Un groupe génique peut subir une expansion ou une contraction par un **crossing-over inégal**, lorsqu'une recombinaison survient entre deux gènes non alléliques (**Figure 19.3**). Habituellement, la recombinaison implique deux séquences d'ADN correspondantes, en alignement exact sur les deux chromosomes homologues. Cependant, lorsqu'il y a deux copies d'un gène sur chaque chromosome, de temps en temps, un mauvais alignement peut survenir pour leur permettre de s'apparier. (Ceci nécessite que certaines des régions flanquantes ne soient pas appariées.)

Lorsqu'une recombinaison survient entre deux copies de gènes non appariés, elle engendre des **chromosomes à recombinaison non réciproque**, dont l'un porte une duplication du gène et l'autre une délétion. Le premier recombinant subit une augmentation du nombre de copies du gène qui passe de 2 à 3, tandis que le second présente une diminution de 2 à 1.

Dans cet exemple, nous avons traité les copies de gène 1 et 2, qui ne sont pas en correspondance, comme si elles étaient entièrement homologues. Cependant, un

crossing-over inégal peut aussi se produire lorsque les gènes adjacents sont très ressemblants (bien que la probabilité soit moindre que lorsqu'ils sont identiques).

Un obstacle au crossing-over inégal réside dans la structure morcelée des gènes. Dans un cas tel que les globines, les exons correspondants aux copies de gènes adjacents sont probablement suffisamment semblables pour s'apparier ; mais les séquences des introns peuvent être considérablement divergentes. Le fait que l'appariement se borne aux exons réduit considérablement la longueur continue d'ADN qui peut être impliquée dans un crossing-over inégal et, de ce fait, en diminue la possibilité. Ainsi les divergences entre les introns pourraient accroître la stabilité des groupes de gènes en empêchant des crossing-over inégaux de se produire.

DE NOMBREUSES THALASSÉMIES SONT DUES À DES CROSSING-OVER INÉGAUX

La **thalassémie** résulte d'une mutation qui réduit ou empêche la synthèse de la globine α ou β . La production de crossing-over inégaux dans les groupes de gènes de la globine humaine est révélée par la nature de certaines thalassémies.

La plupart des thalassémies les plus graves sont le résultat de la délétion d'une partie du groupe du gène de globine impliqué. Au moins dans certains cas, les extrémités de la délétion sont situées dans des régions homologues ; la même chose se produirait exactement si la délétion était engendrée par des crossing-over inégaux.

On trouve deux types principaux de délétion parmi les thalassémies. L' **α -thal-1** élimine les deux gènes α . L' **α -thal-2** élimine seulement l'un des deux gènes α . Le nom des thalassémies se réfère au chromosome particulier qui véhicule la délétion. Selon la combinaison diploïde des chromosomes thalassémiques, un individu peut avoir un nombre de chaînes α allant de zéro à quatre. Il y a peu de différences entre un individu qui a le type « sauvage » (qui est de quatre gènes α) et un individu qui a deux ou trois gènes α . Lorsqu'il n'y a qu'un gène α , l'excès des chaînes β forme un tétramère inhabituel β_4 , qui cause la **maladie HbH**. L'absence totale de gènes α entraîne une anasarque fœto-placentaire (**hydrops fœtal**) fatale avant ou à la naissance.

La **Figure 19.4** résume les délétions qui causent ces thalassémies α . Les délétions de l' **α -thal-1** sont étendues, diversement localisées à l'extrémité gauche, les positions des extrémités droites étant encore inconnues. Les délétions

* N. d. T. : Une région située à l'extrémité 5' du groupe β est indispensable à l'expression de tous les gènes du groupe.

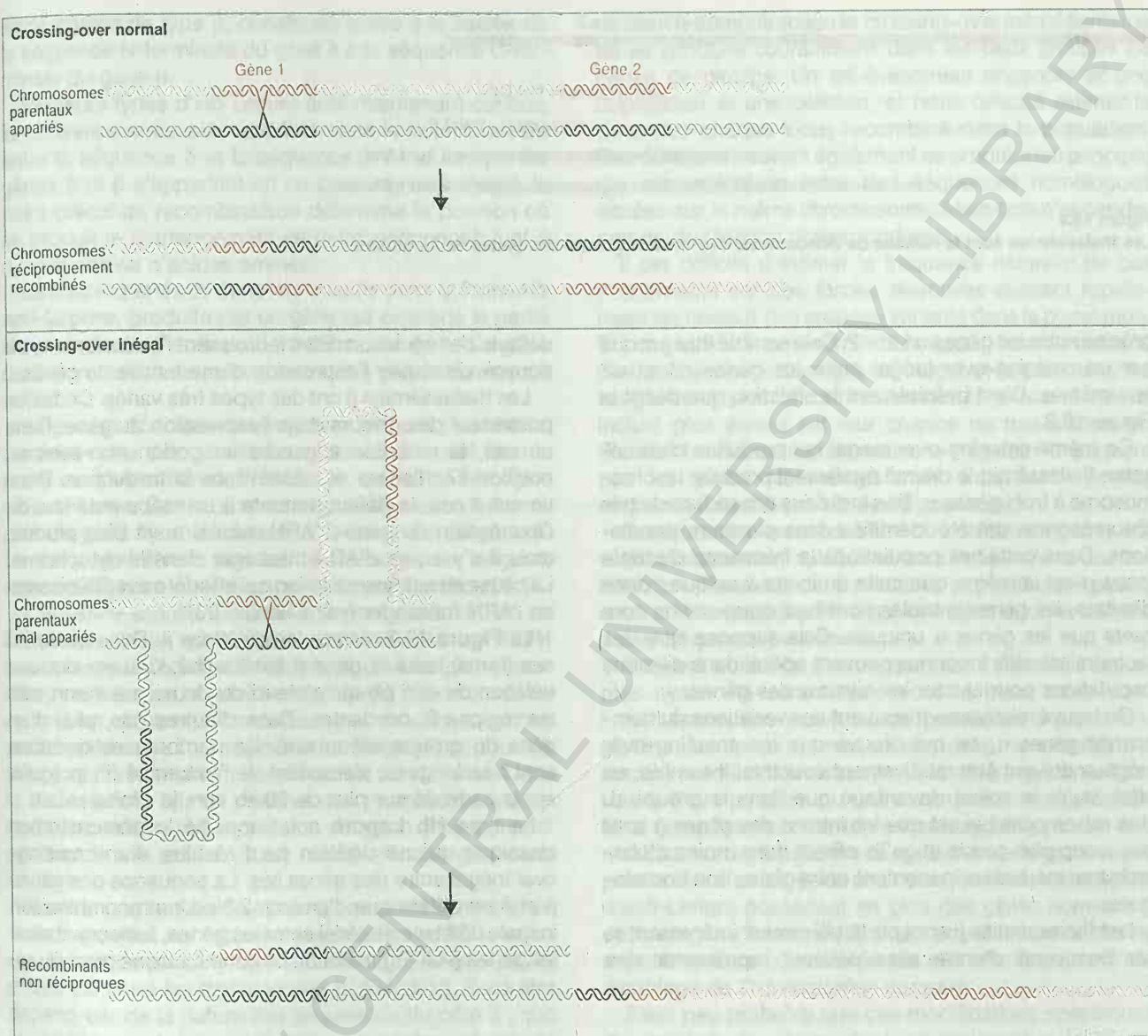


Figure 19.3

Le nombre de gènes peut être modifié par des crossing-over inégaux.

Si le gène 1 d'un chromosome s'apparie avec le gène 2 d'un autre chromosome, les autres copies des gènes ne peuvent s'apparier comme l'indiquent les boucles rejetées à l'extérieur. La recombinaison entre les gènes mal appariés produit un chromosome possédant une copie recombinante unique du gène et un chromosome possédant trois copies du gène (un pour chaque parent et un recombinant).

tions de l' α -thal-2 sont courtes. La forme L élimine 4,2 kb d'ADN comprenant le gène $\alpha 2$. Cela résulte probablement de crossing-over inégal, car les extrémités de la délétion

sont situées dans des régions homologues, respectivement à la droite des gènes $\psi \alpha 1$ et $\alpha 2$. La forme R résulte de l'élimination d'exactly 3,7 kb d'ADN, la distance



Figure 19.4

Les thalassémies sont le résultat de délétions variées dans le groupe des gènes de l' α globine.

précise entre les gènes $\alpha 1$ et $\alpha 2$. Cela semble être produit par un crossing-over inégal entre les gènes $\alpha 1$ et $\alpha 2$ eux-mêmes. C'est précisément la situation que décrit la Figure 19.3.

Ce même crossing-over inégal qui produit le chromosome thalassémique devrait également produire un chromosome à trois gènes α . Des individus possédant de tels chromosomes ont été identifiés dans plusieurs populations. Dans certaines populations la fréquence du triple locus α est la même que celle du locus α unique ; dans d'autres, les gènes α triples sont beaucoup *moins courants* que les gènes α uniques. Cela suppose que des facteurs sélectifs inconnus peuvent opérer dans diverses populations pour ajuster les niveaux des gènes.

On trouve relativement souvent des variations du nombre de gènes α , ce qui prouve que les crossing-over inégaux doivent être relativement courants. Il semble, en effet, qu'ils le soient davantage que dans le groupe β . Une raison possible est que les introns des gènes α sont beaucoup plus courts et qu'ils offrent donc moins d'obstacles au mauvais appariement entre gènes non homologues.

Les thalassémies β sont particulièrement intéressantes car beaucoup d'entre elles peuvent représenter des

défauts de régulation. Elles fournissent l'occasion rare de pouvoir disséquer l'expression d'une famille de gènes.

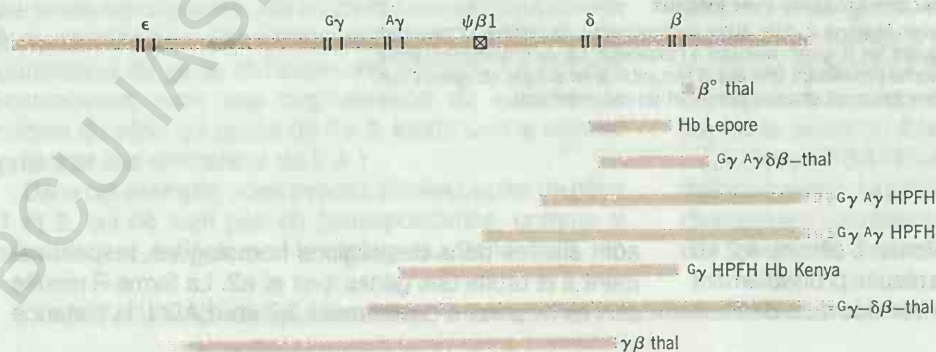
Les thalassémies β ont des types très variés. Certaines possèdent des erreurs dans l'expression du gène. Dans un cas, la mutation engendre un codon non-sens en position 17 ; l'erreur est donc dans la traduction. Dans un autre cas, le défaut remonte à un stade antérieur de l'expression du gène ; l'ARN nucléaire est bien produit, mais il n'y a pas d'ARN messager dans le cytoplasme. La cause en est une mutation qui interfère avec l'épissage de l'ARN messager (voir Chapitre 23).

La Figure 19.5 résume les délétions β . Dans quelques cas (rares), seul le gène β est touché. Ceux-ci ont une délétion de 600 pb qui s'étend du deuxième intron vers les régions 3' bordantes. Dans d'autres cas, plus d'un gène du groupe est touché. De nombreuses délétions sont très longues, s'étendant de l'extrémité 5', indiquée sur la carte, et sur plus de 50 kb vers la droite.

Le type **Hb Lepore** nous apporte la démonstration classique qu'une délétion peut résulter d'un crossing-over inégal entre des gènes liés. La séquence des gènes β et δ ne diffère que d'environ 7 %. Une recombinaison inégale délète le matériel entre les gènes, fusionnant ainsi les gènes (voir Figure 19.3). Le gène fusionné produit une

Figure 19.5

Des délétions dans un groupe de gènes de β globine provoquent plusieurs types de thalassémies.



seule chaîne de type β , constituée grâce à la liaison de la séquence N-terminale du gène δ à la séquence C-terminale du gène β .

Plusieurs types d'Hb Lepore sont maintenant connus, la différence entre eux reposant dans le point frontière entre la séquence δ et la séquence β . Ainsi lorsque les gènes δ et β s'apparient en un crossing-over inégal, le point précis de recombinaison détermine la position où se produit le réarrangement entre les séquences δ et β dans la chaîne d'acides aminés.

La réciproque de ce fait a été trouvée dans la forme **Hb anti-Lepore**, produite par un gène qui possède la partie N-terminale de β et la partie C-terminale de δ . Ce gène de fusion devrait se trouver entre des gènes normaux δ et β .

La preuve que des crossing-over inégaux peuvent se produire entre des gènes moins semblables est fournie par la caractérisation de l'**Hb Kenya**, une autre hémoglobine de fusion. Cette forme contient la séquence N-terminale du gène γ et la séquence C-terminale du gène β . La fusion doit être le résultat de crossing-over inégaux, entre γ et β , dont les séquences diffèrent d'environ 20 %.

L'absence simultanée de δ et β peut produire l'un ou l'autre des deux phénotypes suivants. Dans l'**HPFH** (persistance héréditaire de l'hémoglobine fœtale = hereditary persistence of fetal hemoglobin) il n'existe pas de symptôme clinique ; la maladie a peu de conséquences car la synthèse de l'hémoglobine fœtale ($\alpha_2\gamma_2$) se poursuit même après le moment où elle aurait dû s'arrêter. Dans la **thalassémie $\delta\beta$** , on observe des symptômes anémiques, car même si l'expression du gène γ se poursuit pendant la vie adulte, il est moins efficace que dans l'HPFH.

Il existe nombre de délétions bloquant la synthèse de δ et de β ; elles sont énumérées à la Figure 19.5, selon les classes de phénotypes. Nous ne comprenons pas la différence entre les thalassémies HPFH et $\delta\beta$. Peut-être dépend-elle de la nature des séquences du côté 3', que la délétion met en juxtaposition avec les gènes de globine *.

LES GROUPES DE GÈNES SUBISSENT DE CONTINUELLES RÉORGANISATIONS

Des différences entre les groupes de gènes de globine chez des mammifères différents, nous pouvons déduire que la duplication (parfois) suivie d'une divergence a constitué un fait important dans l'évolution de chacun des groupes. Les délétions thalassémiques humaines

tendent à démontrer que le crossing-over inégal continue de se produire couramment dans les deux groupes de gènes de globine. Un tel événement engendre *et* une duplication *et* une délétion, et nous devons estimer le destin de chaque locus recombiné dans la population. Des délétions peuvent également se produire (en principe) par recombinaison entre des séquences homologues situées sur le *même* chromosome. Mais cela n'engendre pas de duplication correspondante.

Il est difficile d'estimer la fréquence naturelle de ces événements car des forces sélectives ajustent rapidement les niveaux des groupes variants dans la population. Il y a peut-être une corrélation approximative entre la probabilité d'un crossing-over inégal et la parenté entre les gènes — plus ils sont ressemblants (exons et introns inclus) plus élevée est leur chance de mal s'apparier. (Pourtant certaines recombinaisons inégales n'impliquent pas les gènes eux-mêmes mais s'appuient sur des séquences répétitives situées à proximité.)

Généralement une réduction du nombre de gènes est vraisemblablement nuisible et sera l'objet d'une contre-sélection. Pourtant, dans certaines populations, il peut y avoir un avantage équilibré qui maintient la forme délétée à une fréquence peu élevée.

Quel est le résultat d'une expansion ? Les seuls exemples ayant été caractérisés sont le triple locus α et l'anti-Lepore. Les individus qui possèdent 5 gènes d' α globine (un locus normal et un triple) ne manifestent aucune modification de la synthèse d'hémoglobine. Toutefois, il est possible qu'une augmentation supplémentaire (jusqu'au taux de triple homozygotie) puisse être nuisible, en déséquilibrant la synthèse de la globine par production d'un excès de chaînes α . Les individus atteints d'anti-Lepore possèdent en plus des gènes normaux β et δ , un gène fusionné $\beta\delta$. Il est possible que la chaîne supplémentaire soit délétée car elle interfère dans l'assemblage de l'hémoglobine normale.

Il est peu probable que ces modifications spécifiques du nombre de gènes aient un avantage sélectif qui permettrait leur propagation dans la population. Toutefois, les structures des groupes humains actuels montrent plusieurs duplications qui attestent de l'importance de tels mécanismes. Les séquences *fonctionnelles* comprennent deux gènes α codant pour la même protéine, deux gènes β et δ relativement homologues et deux gènes γ pratiquement identiques. Ces duplications indépendantes, relativement récentes, se sont maintenues dans la population, sans parler bien entendu des duplications plus anciennes qui, à l'origine, engendraient les

* N. d. T. : Les délétions plus étendues du côté 5' dans les HPFH que dans les thalassémies $\beta\delta$ pourraient aussi affecter la région du génome responsable de la commutation $Hb\gamma \rightarrow Hb\beta$, qui survient à la naissance.

sent par hasard avec une probabilité (plus ou moins) égale dans toutes les régions du génome, à l'exception de « points chauds » (hot spots) où la fréquence des accidents est très augmentée. La plupart des mutations qui modifient la séquence de l'acide aminé sont *délétères* et seront éliminées relativement rapidement par sélection naturelle (le taux d'élimination dépendant de l'étendue du préjudice).

Peu de mutations seront bénéfiques mais celles qui le sont peuvent se propager dans la population en se substituant éventuellement à la séquence précédente. Lorsqu'un nouveau variant remplace la première version d'un gène, on dit qu'il s'est **établi** dans la population. Il reste un point litigieux concernant la proportion de mutations d'une séquence d'acides aminés pouvant être **neutres**, c'est-à-dire, sans effet aucun sur la fonction de la protéine, et donc capables de s'accumuler comme résultat d'une **dérivée aléatoire suivie d'une implantation**.

Le taux d'accumulation des mutations est caractéristique pour chaque protéine, et dépend probablement, au moins en partie, de sa souplesse d'adaptation par rapport aux modifications. A l'intérieur d'une même espèce, une protéine évolue par substitutions dues à des mutations, puis par élimination ou implantation à l'intérieur d'un groupe donné de fécondité croisée, c'est-à-dire d'une espèce. La présence dans une population de deux variants alléliques ou plus, est appelé **polymorphisme**.

Un polymorphisme peut être stable ; dans ce cas aucune des formes n'a d'avantage sur une autre. Ou bien le polymorphisme peut être transitoire, comme cela doit être le cas lorsqu'un variant en remplace un autre. Quand nous observons de près un groupe de gènes d'une espèce donnée, nous ne voyons que les variants qui ont survécu.

Lorsqu'une espèce se sépare pour former deux nouvelles espèces, chacune constitue alors un groupe d'évolution indépendant. En comparant dans les deux espèces, les protéines correspondantes, nous pouvons voir les différences qui ont été accumulées entre elles *depuis le moment où leurs ancêtres ont cessé de se croiser*. Certaines protéines sont très bien **conservées** et ne montrent qu'un changement léger, ou même aucun, d'une espèce à l'autre. Cela prouve que presque toutes les modifications sont délétères, et donc contre-sélectionnées.

La différence entre deux protéines s'exprime en fonction de leur **divergence** : le pourcentage de résidus qui diffèrent. La divergence entre des protéines peut ne pas être identique à la différence entre des séquences correspondantes d'acides nucléiques. L'origine de cette différence repose sur la représentation de chaque acide

aminé en un codon à trois bases dans lequel la troisième base n'a souvent aucun effet sur la signification.

On peut potentiellement diviser la séquence nucléotidique en **sites de substitution**, et **sites muets**. A l'endroit des sites de substitution, une mutation altère l'acide aminé codé. L'effet de la mutation délétère, neutre ou bénéfique, dépend du résultat de la substitution de l'acide aminé. Dans les sites muets, la mutation ne fait que substituer un codon synonyme à un autre, il n'y a donc aucun changement dans la protéine. Habituellement les sites de substitution constituent 75 % d'une séquence codante et les sites muets 25 %.

En plus de ses séquences codantes, un gène contient des régions non traduites. Là aussi, les mutations sont potentiellement neutres, à l'exception de certains effets sur la structure secondaire ou sur les signaux régulateurs (habituellement courts).

Bien que les mutations silencieuses soient neutres en ce qui concerne la protéine, elles peuvent affecter l'expression des gènes par un changement de la séquence d'ARN. Par exemple, une modification dans la structure secondaire pourrait influencer la transcription, les maturations, ou la traduction. Une autre possibilité est qu'un changement dans des codons synonymes entraîne la réponse d'un ARNt différent influençant ainsi l'efficacité de la traduction.

Les mutations portant sur des sites de substitution devraient correspondre à la divergence de l'acide aminé, ce qui est essentiellement un calcul du pourcentage de modifications. En fait, la divergence ainsi calculée sous-estime les différences qui se sont produites au cours de l'évolution, à cause de la survenue d'événements multiples dans un seul codon. On a donc l'habitude de corriger la divergence. Sans cette correction, la divergence de 0,45 % des acides nucléiques aux sites de substitution correspond à la divergence de 1 % des acides aminés (en supposant que le nombre moyen de sites de substitution par codon est de 2,25).

Pour prendre l'exemple des chaînes de globine humaine β et δ , on note 10 différences dans 146 résidus, soit une divergence de 6,9 %. La séquence d'ADN possède 31 modifications dans 441 résidus. Cependant ces changements sont répartis de façon très différente dans les sites de substitution et les sites muets. Il y a 11 modifications dans les 330 sites de substitution, mais 20 dans seulement 111 sites muets. Cela nous donne (en donnée corrigée) des taux de divergence de 3,7 % dans les sites de substitution et de 32 % dans les lieux muets, presque une différence d'un ordre de grandeur.

La différence frappante entre la divergence des sites de substitution et celle des sites muets démontre l'existence de contraintes beaucoup plus importantes sur les

positions des nucléotides qui influencent la constitution de la protéine que sur les positions qui ne l'influencent pas. Probablement très peu de modifications de l'acide aminé sont neutres.

Si nous prenons le pourcentage de mutations dans les sites muets pour indiquer le pourcentage sous-jacent d'établissement d'une mutation (cela pour montrer qu'il n'y a absolument aucune sélection des sites muets), sur la période couverte depuis que les gènes β et δ ont divergé, il devrait y avoir 32 % de modifications pour 330 sites de substitution, soit un total de 105 sites touchés. Toutes les mutations sauf 11 ont été éliminées, ce qui signifie qu'environ 90 % des mutations n'ont pas survécu.

L'HORLOGE DE L'ÉVOLUTION RETRACE LE DÉVELOPPEMENT DES GÈNES DE LA GLOBINE

Lorsque l'on examine une protéine spécifique dans un type d'espèce, la divergence entre les séquences de chaque couple comparé est (plus ou moins) proportionnelle au temps qui s'est écoulé depuis leur séparation. Cela nous fournit une **horloge d'évolution** qui mesure l'accumulation des mutations selon un taux apparemment aléatoire pendant l'évolution d'une protéine donnée.

Le pourcentage des divergences développées est calculé en pourcentage de la différence par million d'années, ou réciproquement on calcule la période d'évolution par unité (UEP, unit evolutionary period) selon le temps en million d'années qu'il faut pour développer une divergence de 1 %. Une fois établie, l'échelle de temps d'après les comparaisons des couples et selon les espèces (tout en se souvenant des difficultés pratiques pour établir le temps réel d'individualisation des espèces), on peut l'appliquer aux gènes apparentés à l'intérieur d'une espèce. D'après leur divergence on peut donc calculer le temps qui s'est écoulé depuis la duplication dont ils sont originaires.

En comparant les séquences de gènes homologues dans des espèces différentes, il est possible de déterminer le pourcentage de divergence dans les sites muets et de substitution.

En comparant les couples, un taux de divergence de 10 % apparaît dans les sites de substitution des gènes de globine α ou β des mammifères qui se sont séparés lorsque l'individualisation des mammifères s'est produite il y a quelque 85 millions d'années. Cela correspond à un taux de divergence des sites de substitution de 0,12 % par million d'années.

Le pourcentage reste stable lorsque l'on étend la comparaison aux gènes qui ont divergé dans un passé plus éloigné. Par exemple, la divergence moyenne des

sites de substitution entre des gènes de globine de mammifères et de poulet correspondants est de 23 %. Relativement à une séparation datant d'environ 270 millions d'années, cela nous donne un pourcentage de 0,09 % par million d'années.

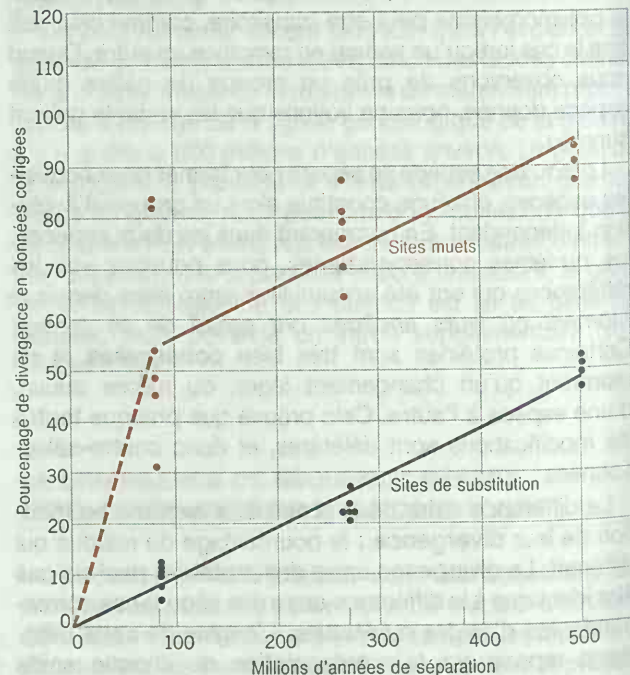
En remontant le temps encore plus avant, on peut comparer le gène de globine α avec le gène de globine β à l'intérieur d'une espèce. Ils ont divergé depuis que les formes individuelles du gène se sont séparées il y a au moins 500 millions d'années (voir la Figure 19.6). Ils ont un taux moyen de divergence des sites de substitution d'environ 50 %, ce qui donne un pourcentage de 0,1 % par million d'années.

Ces données sont reportées sur la **Figure 19.7**, qui montre que le taux moyen de divergence des sites de substitution dans les gènes de globine est d'environ 0,096 % par million d'années (soit une UEP de 10,4). Si l'on considère les incertitudes dans l'estimation des moments précis (dans le temps) où les espèces ont divergé, les résultats obtenus confirment l'idée qu'une échelle linéaire de temps existe bien.

Les données pour les sites muets sont beaucoup moins claires. Quel que soit le cas il est évident que la divergence d'un site muet est beaucoup plus importante que celle d'un site de substitution, différence d'un facteur compris

Figure 19.7

La divergence des séquences d'ADN dépend de la séparation évolutive. Chaque point du graphique représente la comparaison d'un couple.



entre 2 et 10. Mais l'extension des divergences des sites muets, en comparaison par couples, est trop importante pour indiquer si une échelle linéaire de temps est applicable (nous devons donc baser nos comparaisons dans le temps sur les sites de substitution).

D'après la Figure 19.7, il est clair que le pourcentage par rapport au temps dans les sites muets n'est pas linéaire. *Si nous considérons qu'il doit exister une divergence zéro à zéro année de séparation*, nous voyons que le pourcentage de divergence d'un site muet est beaucoup plus important pendant les 100 premiers millions d'années de séparation. Il est possible qu'une fraction correspondant à environ une moitié des sites muets soit rapidement saturée par les mutations (dans les 100 premiers millions d'années); cette moitié se conduit comme des sites neutres. L'autre moitié accumule les mutations plus lentement, à un taux qui est approximativement le même que celui des sites de substitution; cette fraction caractérise des sites silencieux à l'égard de la protéine mais qui, pour une autre raison, ont été soumis à une pression sélective.

On peut maintenant inverser le calcul des pourcentages de divergence pour estimer les périodes où les gènes d'une espèce se sont séparés. La différence entre les gènes humains β et δ est de 3,7 % pour les sites de substitution. A une UEP de 10,4, ces gènes doivent avoir divergé il y a $10,4 \times 3,7 \approx 40$ millions d'années, à peu près au moment de la séparation des branches conduisant aux primates du Nouveau Monde, aux primates de l'Ancien Continent, aux grands singes et aux humains. Tous ces primates supérieurs possèdent à la fois les gènes β et δ , ce qui suggère que la divergence du gène a commencé juste avant ce point de l'évolution.

Si l'on remonte encore plus loin, la divergence entre les sites de substitution des gènes γ et ϵ est de 10 %, ce qui correspond à une séparation remontant à environ 100 millions d'années. La séparation entre les gènes de globine embryonnaires et fœtaux doit dès lors avoir juste précédé ou même accompagné la divergence des mammifères.

Un arbre concernant l'évolution des gènes de globine humains est construit à la Figure 19.8. On devrait trouver chez tous les mammifères des caractéristiques ayant évolué avant la séparation de la branche des mammifères — telle que la séparation des gènes de type β adultes et pré-adultes. Quant aux caractéristiques qui ont évolué par la suite — telle que la séparation des gènes de globine β et δ — elles ont dû le faire indépendamment dans des lignées individuelles de mammifères.

Dans chaque espèce il y a eu des modifications qui, en comparaison, apparaissent récentes de la structure des groupes, puisque nous constatons des différences

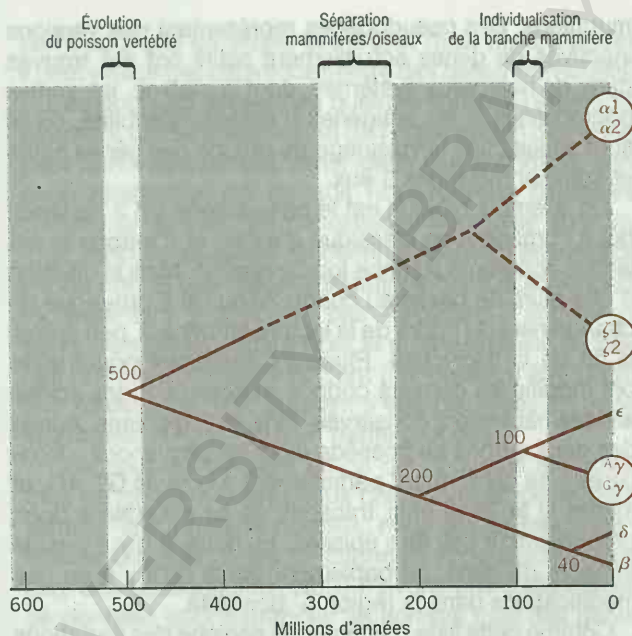


Figure 19.8

La divergence des sites de substitution entre des couples de gènes de β globine nous permettent de reconstruire l'histoire du groupe humain.

Cet arbre rend compte de la séparation des classes de gènes de globine. L'origine des duplications de gènes individuels n'est pas connue, comme le montrent les cercles. On ignore le moment de la divergence entre α et ζ .

dans le nombre de gènes (un gène de β globine chez l'homme, deux chez la souris) ou dans le type de gènes (nous ne sommes pas encore sûrs que les globines de type β embryonnaires et fœtales soient séparées chez le lapin et la souris).

LES PSEUDOGÈNES SONT LES « VOIES SANS ISSUE » DE L'ÉVOLUTION

Les pseudogènes sont définis comme tels parce qu'ils possèdent des séquences ressemblant à celles des gènes fonctionnels, mais qui ne peuvent être traduites en une protéine fonctionnelle.

Certains pseudogènes ont la même structure générale que les gènes fonctionnels, avec des séquences correspondant aux exons et aux introns à la place habituelle. Ils sont rendus inactifs par des mutations qui bloquent un ou tous les stades de l'expression du gène. Ces altérations peuvent être soit une abolition des signaux d'initiation de la transcription, soit un blocage de l'épissage aux jonctions exon-intron, soit la fin prématurée de la traduction.

Habituellement, un pseudogène possède plusieurs mutations délétères, probablement parce qu'ayant cessé d'être actif, rien n'empêche plus l'accumulation des

mutations. Des pseudogènes représentant des versions inactives de gènes actuellement actifs ont été trouvés dans de nombreux systèmes, dont la globine, les immunoglobulines et les antigènes d'histocompatibilité, où ils sont situés dans le voisinage du groupe des gènes actifs et souvent mélangés à eux.

Un exemple typique est le pseudogène $\psi\beta 2$ du lapin, qui a l'organisation classique d'exons et d'introns et est le plus apparenté au gène fonctionnel $\beta 1$. Mais la délétion d'une paire de bases au codon 20 du $\psi\beta 2$ provoque un changement du cadre de la lecture entraînant, peu après, la fin de la traduction. Plusieurs mutations ponctuelles ont modifié les derniers codons représentant des acides aminés hautement conservés dans les β globines. Aucun des deux introns ne possède alors les séquences « frontières » se conformant parfaitement à la règle GT-AG, et même si le gène était transcrit, ils ne pourraient donc probablement pas être épissés. En outre, il n'y a pas de transcrit du gène, probablement parce qu'il y a eu des modifications dans la région 5' bordante.

Comme cette liste d'altérations englobe des mutations bloquant potentiellement chaque stade de l'expression du gène, nous n'avons aucun moyen de dire quel événement a rendu le gène inactif à l'origine. Cependant, d'après la divergence entre le pseudogène et le gène fonctionnel, nous pouvons estimer à quel moment est apparu le gène et à quel moment ses mutations ont commencé à s'accumuler.

Si le pseudogène était devenu inactif dès sa formation par duplication du gène $\beta 1$, les pourcentages de divergence des sites de substitution et des sites muets devraient être les mêmes. (Il n'y a pas de raison qu'ils soient différents si le gène n'est pas traduit.) Mais en réalité, il y a moins de mutations des sites de substitution que des sites muets. Cela suggère qu'il y a tout d'abord eu une contre-sélection des mutations aux sites de substitution. D'après l'importance respective des mutations dans les deux types de sites, on peut calculer que $\beta 2$ a divergé de $\beta 1$ il y a environ 55 millions d'années, qu'il est resté un gène fonctionnel pendant 22 millions d'années, mais qu'il est un pseudogène depuis les dernières 33 millions d'années.

Des calculs similaires peuvent être réalisés pour d'autres pseudogènes. Certains semblent également avoir été actifs pendant un certain temps avant de devenir des pseudogènes. D'autres encore semblent avoir été inactifs dès la génération originelle. Le point commun aux structures de ces pseudogènes est que chacun d'entre eux a évolué indépendamment pendant le développement des

groupes de gènes de globine de chaque espèce. Cela renforce la conclusion selon laquelle la création de nouveaux gènes suivie de leur reconnaissance en tant que réplique fonctionnelle, les remaniements les transformant en nouveaux gènes fonctionnels ou leur inactivation sous forme de pseudogènes, est un processus continu dans le groupe génique.

Le gène $\psi\alpha 3$ de la souris a une propriété intéressante : en effet, il ne possède aucun des deux introns. On peut aligner sa séquence (autorisant les mutations accumulées) avec celle de l'ARN messager de l' α globine. Le moment apparent d'inactivation coïncide avec la duplication originelle, ce qui suggère que l'événement inactivant originel peut avoir été associé à la perte des introns. Sans introns, le gène aurait été incapable de fonctionner et donc aurait immédiatement commencé à accumuler des mutations.

Comment les introns ont-ils pu être perdus ? Il n'y a aucune raison pour que les systèmes de reconstruction de l'ADN aient reconnu les frontières exon-intron, ce qui suggère qu'à une certaine étape il y a eu implication de l'ARNm lui-même. Il y a probablement eu insertion d'une copie inverse de l'ARNm dans le génome, peut-être véhiculée par un rétrovirus pour lequel une telle action est normale (voir Chapitre 30). Le plus difficile à imaginer serait un mécanisme selon lequel il y aurait appariement entre l'ARN messager et un brin d'ADN à un moment donné du processus de recombinaison.

Les séquences génomiques inactives qui ressemblent à des copies d'ARN sont appelées des **rétrogènes** (**processed pseudogenes**). Si l'on admet l'idée qu'ils ont trouvé leur origine dans l'insertion d'un produit dérivé de l'ARN en un site aléatoire, ces pseudogènes peuvent se trouver n'importe où dans le génome, et même pas nécessairement sur le même chromosome que le gène actif.

Les pseudogènes sont-ils courants ? La plupart des familles de gènes comportent des membres pseudogènes. En général les pseudogènes représentent une petite minorité du nombre total des gènes. Cependant il y a un cas exceptionnel où un seul gène actif code pour une protéine ribosomale de souris et possède environ 15 « rétrogènes » ; ceci doit être pris en considération lorsque nous essayons de calculer un nombre de gènes d'après les résultats d'une hybridation*.

Si les pseudogènes sont les « voies sans issue » de l'évolution et ne sont que les corollaires indésirables du réarrangement des gènes fonctionnels, pourquoi sont-ils encore présents dans le génome ? Remplissent-ils une

* N. d. T. : En fait plusieurs exemples ont été décrits où existent un seul gène actif et de très nombreux pseudogènes : c'est le cas par exemple du gène de la glyceraldéhyde phosphate déshydrogénase, de l'aldolase A chez le rat et de l'argininosuccinate synthase. Dans certains de ces exemples, il existe une cinquantaine de pseudogènes pour un seul gène actif.

fonction ou bien sont-ils entièrement sans objet, auquel cas il faudrait pour qu'ils se maintiennent que la pression sélective ne s'exerçât point.

Nous devons garder à l'esprit que nous avons vu de tels gènes survivre dans des populations actuelles. Dans le passé un nombre incalculable d'autres pseudogènes peut avoir été éliminé. Cette élimination pourrait se produire par brusque délétion de la séquence ou par accroissement des mutations à un point tel que le pseudogène ne pourrait plus être reconnu comme membre de sa famille de séquence originelle (c'est probablement le destin d'un pseudogène qui n'a pas été brusquement éliminé).

Même des reliques de l'évolution peuvent être dupliquées. Dans les gènes de β globine de chèvre, existent deux espèces adultes β^A et β^C . Chacune des deux possède un pseudogène éloigné de quelques kilobases en amont (appelés respectivement $\psi\beta^Z$ et $\psi\beta^X$). Les deux pseudogènes se ressemblent plus entre eux qu'ils ne ressemblent aux gènes adultes de β globine ; en particulier, ils ont en commun plusieurs mutations inactivantes. De même les deux gènes adultes de β globine sont plus semblables entre eux qu'ils ne le sont aux pseudogènes. Cela implique qu'une structure originelle $\psi\beta - \beta$ a elle-même été dupliquée produisant deux gènes fonctionnels β (qui divergent ensuite en donnant les gènes β^A et β^C) et deux gènes non-fonctionnels (qui divergent en donnant les deux pseudogènes actuels).

Les mécanismes responsables de la duplication, de la délétion et du réarrangement d'un gène agissent sur toutes les séquences qui sont reconnues comme membres du groupe, qu'ils soient ou non fonctionnels. Il est laissé à la sélection le rôle de discrimination entre les produits.

DE NOMBREUSES PROTÉINES ONT EN COMMUN DES FAMILLES DE GÈNES

Lors d'expériences visant à évaluer le nombre de gènes par isolement de séquences génomiques correspondant à une sonde spécifique d'ARN messager ou d'ADN complémentaire, on trouve souvent des fragments multiples qui correspondent à la sonde. Bien entendu, avant que leurs séquences nucléotidiques ne soient déterminées, nous ignorons combien sont fonctionnels et combien sont des pseudogènes. En général, il est établi que la plupart des gènes sont actifs. D'après ce critère il y aurait des gènes multiples pour de nombreuses, peut-être pour la plupart des protéines de structure. Les gènes multiples peuvent être groupés ou dispersés. Dans certaines familles, tous les membres partagent une organisa-

tion commune en exons et introns, mais dans d'autres, il peut y avoir des différences entre des gènes individuels.

Ces données relèvent en quelque sorte de l'anecdote, montrant que la préservation de l'organisation des gènes peut considérablement varier dans son ensemble, sans nous renseigner sur ce que nous devons considérer comme la norme. Les conclusions générales qu'il faut tirer, sont que les gènes fonctionnels peuvent être morcelés ou non et que des modifications dans le schéma du morcellement n'affectent pas forcément l'activité du gène. Des gènes codant pour la même protéine ou pour des protéines semblables n'ont pas besoin d'être disposés en tandem, mais peuvent s'être dispersés en gènes individuels ou peut-être en petits groupes.

Que peut-on conclure en ce qui concerne le nombre total de gènes ? Quand nous cherchons un gène représentant une protéine donnée, nous avons tendance à trouver d'autres gènes fonctionnels apparentés et (parfois) des pseudogènes. Nous n'avons pas suffisamment de données pour extrapoler au nombre total des gènes, mais nous pourrions nous demander jusqu'où ce type de répétition peut aller pour apporter une explication à la quantité apparemment excessive d'ADN.

La répétition des séquences augmente la quantité d'ADN impliquée dans une fonction spécifique. Plusieurs gènes fonctionnels peuvent coder pour la même protéine, ou pour des variants inattendus. Lorsque les copies sont groupées, les régions inter-géniques peuvent être relativement importantes, accroissant ainsi la quantité totale d'ADN chargée de déterminer la fonction. Les pseudogènes peuvent être pris en compte dans des séquences supplémentaires en rapport avec la fonction, mais ils ne peuvent plus contribuer à cette fonction.

Il ne serait pas étonnant que la quantité d'ADN impliquée dans une fonction donnée soit 3 ou 4 fois plus importante qu'elle le serait si un gène unique codait pour chaque fonction. En analysant le nombre de gènes au Chapitre 17, nous prédisions un chiffre d'environ 30 000 à 40 000 fonctions géniques chez un mammifère. Si nous considérons à présent que chaque fonction d'un gène est représentée par 3 ou 4 copies (soit apparentées, soit réellement identiques) nous pourrions prédire un nombre total de gènes allant jusqu'à quelque 125 000.

La première hypothèse, lorsque l'on prend en compte la partie codante du génome, est que la plupart des fonctions, contrairement à ce qu'il semblait à première vue, sont représentées par un petit nombre de gènes au lieu de l'être par un gène unique. La seconde hypothèse serait qu'une bonne partie de l'ADN est gaspillée soit à l'intérieur du gène (sous la forme d'introns), soit entre les gènes (parce que les groupes géniques peuvent être éparpillés et qu'il est donc probable qu'ils soient éloignés

les uns des autres). Cette quantité d'ADN extrêmement grande n'est impliquée dans les fonctions des gènes que de façon marginale.

Nous avons fait un progrès tangible dans l'estimation du contenu de l'ADN avec cette (simple) notion qu'il y a davantage de gènes que prévu : lorsque nous caractérisons une fonction spécifique, nous avons souvent affaire à des petites familles plutôt qu'à des gènes isolés. Grâce à cet argument, nous pouvons considérer qu'environ 2×10^8 pb d'un génome mammifère (environ 10 % de l'ADN) peuvent être directement impliqués dans le codage pour des protéines.

Supposons que l'ADN excédentaire contenu à l'intérieur de ces gènes ou dans des séquences intergéniques avoisinantes représente environ 3 fois plus d'ADN, soit encore environ 6×10^8 pb. L'attribution à l'ADN de certaines fonctions pourrait être sous-estimée si les gènes étaient plus grands ou si la distance entre eux était plus importante que supposé. Les 125 000 gènes dont nous avons fait l'hypothèse pourraient-ils prendre en charge la totalité des 3×10^9 pb du génome mammifère ?

Nous avons organisé la majeure partie du génome par la supposition qu'il est associé à des régions codantes ; bien qu'il ne code pas lui-même pour des protéines, il ne possède aucune autre fonction apparente. (Ces régions pourraient contenir de courts éléments régulateurs, par exemple des origines de réplication ou des sites d'identification de la structure chromosomique ; mais il est improbable qu'ils occupent une proportion substantielle de l'ADN global.) A moins que nous ne démontrions que la quantité totale de l'ADN dans le noyau joue un quelconque rôle, la majorité des séquences associées aux régions codantes semble être superflue.

Même si nous avons réussi à expliquer le contenu en ADN du génome de mammifère, nous n'avons pas pris en compte le paradoxe de la valeur de C qui fait que le contenu total de l'ADN peut varier de manière beaucoup plus considérable que le nombre de gènes. Cela signifie-t-il que la taille du gène lui-même, ou que la distance entre les gènes est relativement peu importante pour l'organisme et peut donc être soumise à des réajustements opportunistes ?

CERTAINS GROUPES DE GÈNES EN TANDEM CODENT POUR LES HISTONES

Quelques catégories de gènes sont organisées en grands groupes contenant de nombreuses copies d'un gène (ou de gènes). Les gènes (ou autres séquences) qui sont sous

forme de multiples copies adjacentes sont dits **répétés en tandem**. Une catégorie de groupe de gènes en tandem code pour les protéines histones, une autre pour les ARN ribosomiaux. Dans les deux cas, une abondante répétition peut refléter le besoin de fournir de grandes quantités de produit.

Le groupement en tandem de gènes, dans une région connue dont le but unique est de coder pour une fonction spécifique, nous suggère une façon de décrire l'organisation des séquences. Chaque tandem d'un groupe peut être décrit comme une **unité de répétition**, unité de base dont les répétitions, bout à bout, constituent le groupe. Les copies de l'unité de répétition peuvent être identiques, ou varier. L'unité de répétition est en principe plus longue que la séquence transcrite ; il est donc possible de la diviser en deux types de régions. L'**unité de transcription** est située entre la première et la dernière base transcrite en ARN. L'**intervalle non transcrit** est la région située entre des unités de transcription adjacentes.

Ce modèle d'organisation est illustré par la **Figure 19.9**, qui montre que la carte de restriction d'un groupe de gènes en tandem possède des traits caractéristiques. Elle est circulaire. Lorsque les unités de répétition sont identiques, chacune d'entre elles engendre le même schéma de fragments de restriction. Des répétitions adjacentes ont en commun les mêmes fragments terminaux. Dans l'exemple montré par la figure, le fragment A est adjacent au fragment B, qui est adjacent au fragment C, qui à son tour est adjacent au fragment A, ce qui produit ainsi une carte circulaire.

Ce schéma n'est interrompu qu'aux extrémités du groupe, représentées dans la figure par les fragments X et Y. Mais généralement les fragments des extrémités ne sont pas détectés, parce que chacun est unique, alors que les fragments internes sont répétés. Par exemple, s'il y avait 100 copies de l'unité de répétition, les fragments X et Y ne constitueraient que 1 % du matériel présent dans les fragments A, B et C.

L'intervalle non transcrit est habituellement plus court que l'unité de transcription. Nous pouvons caractériser cet intervalle grâce à la clarté avec laquelle les régions sont délimitées dans un groupe de gènes en tandem. L'intervalle constitue simplement une partie de l'unité de répétition non exprimée dans l'ARN mais dont les fonctions (s'il en a) restent centrées sur le but unique du groupe : la production de l'ARN codé par l'unité de transcription.

Cette vision nous pousse à nous demander s'il est possible de considérer sous le même angle, des séquences plus longues se situant entre chaque gène (par exemple) du groupe des gènes de globine. Il y a cepen-

Organisation linéaire du groupe

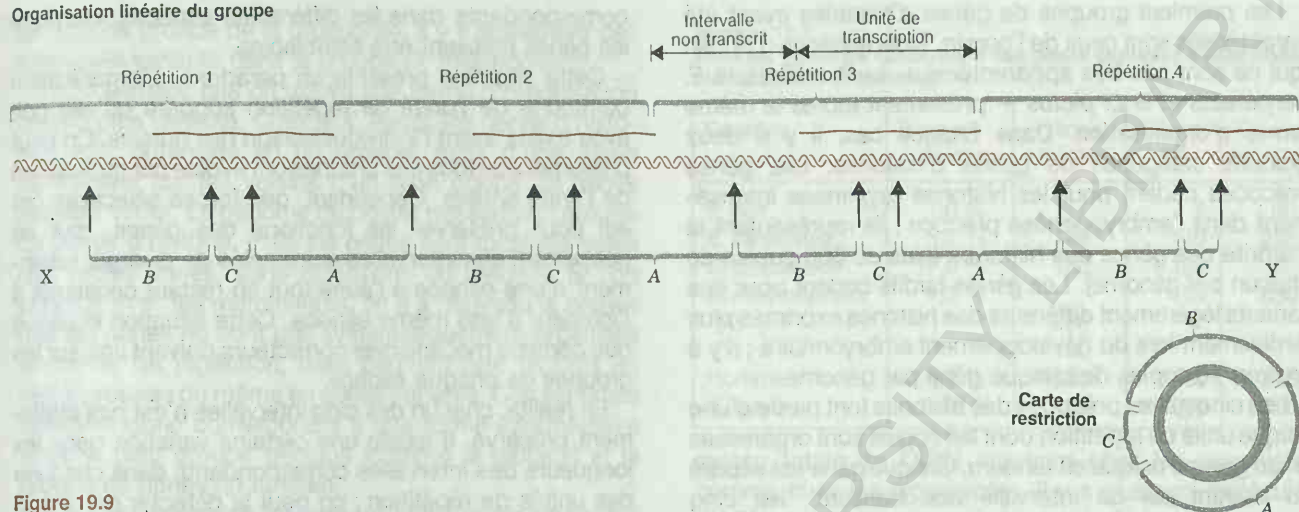


Figure 19.9

Un groupement de gènes en tandem présente une alternance d'unités de transcription et d'intervalles non transcrits ; il produit une carte de restriction circulaire.

dant une grande différence entre les deux situations. Chaque région intergénique d'un groupe de gènes de globine est différente, tandis que les intervalles non transcrits d'un groupe de gènes en tandem sont répétés tout au long des gènes.

Un groupe de gènes en tandem ne code pas nécessairement pour un produit unique ; l'unité de répétition peut contenir plus d'une unité de transcription et plus d'un intervalle non transcrit. Les groupes de gènes des histones nous fournissent l'exemple d'une telle organisation.

Les histones sont les protéines nucléaires de structure *par excellence*. Chez (probablement) tous les eucaryotes, les chromosomes contiennent 5 types de protéines histones : H1 présente en quantité semi-molaire et H2A, H2B, H3 et H4 présentes en quantités équimolaires (voir Chapitre 26). La masse totale des histones est à peu près équivalente à la masse de l'ADN lui-même.

Dans des cellules somatiques en division, la synthèse des histones se produit *en parallèle* à la réplication de l'ADN ; dès qu'un fragment d'ADN est répliqué les histones sont disponibles pour s'associer à lui. Pour que cela soit réalisable, une grande quantité de protéines d'histone doit être synthétisée pendant la courte période où l'ADN est synthétisé. Le besoin pour la cellule somatique de synthétiser la valeur en histones d'un génome, et cela pendant un court laps de temps, pourrait être la raison fondamentale de la répétition du gène d'histone.

Il n'y a pas de rapport particulier entre la taille du génome et le nombre de gènes histones (l'efficacité de l'expression du gène d'histone peut donc varier d'une espèce à l'autre). La répétition de ces gènes est un aspect

courant de l'arrangement nécessaire à la production des histones. En principe il y a le même nombre de copies de chaque gène d'histone. Chez le poulet la fréquence de répétition est à peu près de 10, chez les mammifères elle est environ de 20. Elle s'élève à environ 40 chez *X. laevis* et jusqu'à environ 100 chez *D. melanogaster*. Plusieurs espèces d'oursin possèdent entre 300 et 600 copies de chaque gène d'histone.

Chez les oursins, les divisions nucléaires initiales de l'embryogenèse se produisent en effet très rapidement. Ce grand nombre de gènes chez l'oursin doit être nécessaire pour permettre à l'expression *de novo* du gène d'histone de maintenir à un niveau élevé la synthèse de l'ADN. Chez les amphibiens, l'expression des gènes des histones précède le début de l'embryogenèse, et une grande quantité d'ARN messager des histones s'accumule dans l'oocyte. Quelques protéines histones peuvent être également stockées. Cela apporte une exception à la règle qui veut que les gènes d'histones ne soient exprimés que pendant la phase de synthèse de l'ADN. L'utilisation de différentes méthodes pour fournir des histones en quantité nécessaire explique probablement le manque de relation entre le nombre des gènes d'histone et la valeur de C.

L'organisation des gènes d'histone est un peu variable. Chez les espèces à basse répétition (moins de 50) il n'y a pas de forme d'organisation logique. Chez les espèces à haute répétition (plus de 100), une grande proportion des gènes peut être organisée en une seule sorte de groupe de gènes répétés en tandem. Dans tous les cas, les gènes fonctionnels de l'histone n'ont pas d'introns.

Les premiers groupes de gènes d'histones ayant été caractérisés sont ceux de l'oursin. Trois espèces d'oursin (qui ne sont pas très apparentées) — le *P. miliaris*, le *S. purpuratus* et le *L. pictus* — présentent toutes la même forme d'organisation. Dans chaque cas, il y a deux grandes catégories de gènes d'histones. Les gènes précoces codent pour les histones exprimées intensément dans l'embryogenèse précoce ; ils représentent la majorité des gènes des histones (plus de 300 copies de chacun par génome). Les gènes tardifs codent pour des variants légèrement différents des histones exprimés plus tardivement lors du développement embryonnaire ; il y a environ 10 copies de chaque gène par génome.

Les cinq gènes précoces des histones font partie d'une unique unité de répétition dont les copies sont organisées en un groupe unique en tandem. Chaque gène est séparé du suivant par un intervalle non transcrit ; les cinq intervalles sont différents. Les copies multiples de l'unité de répétition sont virtuellement identiques.

Tous les gènes sont transcrits dans la même direction bien que chacun, indépendamment, produise son propre ARNm. L'organisation prédominante du groupe précoce chez l'oursin est résumée à la **Figure 19.10**. Il montre que la principale différence entre les espèces réside dans les intervalles non transcrits. Ces derniers diffèrent non seulement en longueur mais également en séquence. Il n'y a pas d'hybridation croisée entre les intervalles

correspondants dans les différentes espèces, bien que les gènes puissent être semblables.

Cette situation présente un paradoxe. L'organisation commune de l'unité de répétition suggère qu'elle doit avoir existé avant l'individualisation des oursins. On peut présumer que tous ces groupes ont évolué par duplication de l'unité entière. Cependant, des forces sélectives ont agi pour préserver les fonctions des gènes, tout en permettant aux intervalles entremêlés de diverger totalement d'une espèce à l'autre tout en restant constants à l'intérieur d'une même espèce. Cette situation implique que certains mécanismes correcteurs doivent agir sur les groupes de chaque espèce.

En réalité, chacun des cinq intervalles n'est pas totalement préservé. Il existe une certaine variation dans les longueurs des intervalles correspondants dans chacune des unités de répétition ; on peut la détecter grâce à la (légère) hétérogénéité dans les longueurs des fragments de restriction appropriés. Ces variations entre des unités de répétition sont appelées **micro-hétérogénéités**.

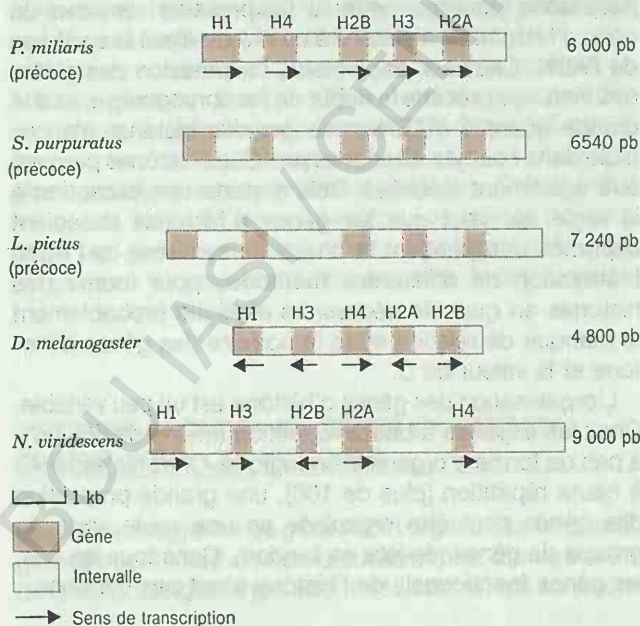
Chez un oursin, deux types d'unités de répétition caractérisant les gènes embryonnaires précoces ont été définis. Ils diffèrent par leurs intervalles. Chaque type d'unité semble être répété en tandem ; les deux sortes de groupes ne sont pas entremêlés. Cela implique que, quels que soient les mécanismes responsables du maintien de l'homogénéité, ils doivent être capables d'agir sur des sous-régions du groupe sans avoir besoin d'agir sur sa totalité.

Les gènes tardifs des histones sont organisés d'une manière différente. On peut trouver ces gènes en fragments individuels d'ADN ; parfois ils sont présents sous la forme d'un couple de gènes faiblement liés (séparés de plus de 10 kb). Il n'y a pas de forme logique d'organisation.

À l'intérieur d'un même organisme, la famille des gènes d'histones peut dès lors être organisée de deux manières relativement différentes : les gènes précoces s'organisent en un grand groupe distinct pendant que les gènes tardifs sont dispersés. Les gènes dispersés proviennent probablement d'une translocation de membres individuels provenant d'un groupe ancestral ; ceci, les libérant des contraintes sélectives visant à maintenir l'homogénéité de la séquence, leur permet ainsi de diverger pour acquérir les fonctions désormais associées aux gènes tardifs.

Chez *D. melanogaster*, les cinq gènes d'histones sont organisés en une unité de répétition unique, bien qu'ici l'organisation de l'unité de répétition ne soit pas la même. Les gènes sont disposés dans un ordre différent et sont transcrits dans des directions différentes. Il faudrait deux translocations pour convertir le type de groupe de l'oursin

Figure 19.10
Les unités de répétition des groupes de gènes d'histones ont des organisations variées.



en type de groupe de la drosophile. Il y a également des intervalles variants dans le groupe *D. melanogaster*.

Dans certains cas on trouve une forme d'organisation encore plus variable. Chez le *X. laevis*, les gènes d'histones sont organisés en un groupe, mais d'une façon hétérogène car les voisins d'un type donné de gènes peuvent varier. Le génome du poulet possède un groupe de gènes d'histone, mais les gènes sont localisés selon des ordres variés, sans répétitions en tandem. Si l'on poursuit avec les mammifères, on a à nouveau la preuve que les gènes ne sont pas systématiquement organisés en unités de répétition, mais peuvent être répartis en petits groupes ou même en gènes séparés. Ces schémas semblent donc correspondre à une forme intermédiaire entre le petit groupe génique (comme la globine) et le groupe de gènes en tandem.

Il y a un contraste intéressant entre la structure des gènes d'histone pris séparément et leur organisation globale. Tous les gènes d'histones partagent la même structure générale ininterrompue *, les gènes correspondants, dans les différentes espèces codent pour des protéines bien apparentées. Il existe cependant une notable flexibilité dans la relation entre les nombreuses classes de gènes, qui varient du groupe arrangé en tandems rigoureux à une disposition apparemment anarchique.

Tout cela renforce les conclusions générales suggérées par l'exemple du groupe des gènes de globine. Les gènes individuels doivent, pour s'adapter à leurs fonctions, être façonnés par des mutations particulières dans des séquences sinon conservées, mais il y a une production de nouvelles copies, et par voie de conséquence un réarrangement de l'organisation globale du groupe.

L'ARN DE TRANSFERT ET L'ARN RIBOSOMAL POSSÈDENT DES GÈNES RÉPÉTÉS

Chacun des composants de l'ARN de l'appareil de synthèse protéique est codé par de nombreux gènes. En un sens, les gènes de structure qui codent pour l'ARN messager forment le plus grand groupe, bien que, du fait qu'il existe une grande diversité d'ARN messagers, chaque gène ne soit représenté que par un nombre relativement réduit de molécules messagères. Pour s'adapter au grand nombre d'ARN messagers qui doivent être traduits, les cellules fabriquent quantité de ribosomes et de molécules d'ARN de transfert.

L'ARN ribosomal est, de loin, le produit prédominant de la transcription ; il constitue quelque 80-90 % de la

Tableau 19.2
Tous les génomes contiennent des gènes multiples d'ARN ribosomal et d'ARN de transfert.

Espèces	Rapport ADNr/total	Gènes 18S/28S	Gènes 5S	Gènes de l'ARNt
<i>E. coli</i>	1,0 %	7	7	60
<i>S. cerevisiae</i>	5,5 %	140	140	250
<i>D. discoideum</i>	17,0 %	180	180	?
<i>D. melanogaster</i>	1,3 %		165	850
(X)		250		
(Y)		150		
Homme	0,4 %	280	2 000	1 300
<i>X. laevis</i>	0,18 %	450	24 000	1 150

masse totale de l'ARN cellulaire chez les eucaryotes comme chez les procaryotes. Il y a plusieurs molécules d'ARNt par ribosome, mais bien sûr elles sont beaucoup plus petites que les ARNr. L'ARNr et l'ARNt sont représentés par des gènes multiples. Le **Tableau 19.2** résume les quantités présentes dans certains génomes.

Le nombre des gènes principaux de l'ARNr varie de 7 chez *E. coli*, à environ 100 à 200 chez les eucaryotes inférieurs, jusqu'à plusieurs centaines chez les eucaryotes supérieurs. Dans pratiquement chaque cas, les gènes des grands et des petits ARNr forment des paires en tandem. (L'unique exception est la mitochondrie de levure.) Dans la bactérie et chez certains eucaryotes inférieurs, le gène codant pour l'ARN 5S fait partie de la même unité, si bien que le nombre total de gènes 5S est le même que celui des ARNr principaux. Chez les bactéries, le gène 5S est co-transcrit avec les gènes principaux d'ARNr ; chez les eucaryotes il est transcrit indépendamment. Chez les eucaryotes supérieurs, les gènes codant pour l'ARN 5S sont organisés séparément dans leurs propres groupes et leur nombre dépasse celui des principaux gènes de l'ARNr.

L'absence de toute variation détectable dans les séquences des molécules d'ARN implique que toutes les copies de chacun des gènes doivent être identiques, ou tout au moins doivent avoir des différences inférieures au seuil de détection dans l'ARNm (qui est d'environ 1 %). Le point d'intérêt capital consiste à déterminer quel(s) mécanisme(s) est (sont) utilisé(s) pour empêcher les variations de s'accumuler dans les séquences individuelles.

Chez les bactéries les paires de gènes multiples 16S-23S de l'ARNr sont dispersées. Dans la plupart des noyaux des eucaryotes, les gènes d'ARNr sont contenus dans un ou plusieurs groupes de gènes en tandem. Ces régions sont parfois appelées **ADNr**. (Dans certains cas la proportion d'ADNr dans l'ADN global, de même que

* N. d. T. : Assez récemment on a découvert un gène d'histone avec intron et de plus, contrairement aux autres gènes, codant pour un messager polyadénylé : il s'agit de l'histone H₃ du poulet dont l'équivalent chez les mammifères semble être l'histone H1₀.

sa composition atypique en bases, est assez importante pour permettre son propre isolement en une fraction séparée, obtenue directement à partir de l'ADN génomique préalablement fragmenté).

Dans le cas de l'ARNt, le nombre exact de gènes est difficile à déterminer, car l'ample structure secondaire de la molécule crée des difficultés techniques dans la réaction d'hybridation. Nous en avons probablement sous-estimé le nombre réel.

LES DEUX GÈNES DE L'ARNr SONT CONTENUS DANS UNE UNITÉ DE RÉPÉTITION EN TANDEM

Les deux ARNr principaux sont transcrits en un précurseur unique chez les bactéries et les noyaux des eucaryotes. Après la transcription, le précurseur est coupé pour libérer les molécules individuelles d'ARNr par des mécanismes que nous discuterons plus en détail au Chapitre 22. Les caractéristiques déterminantes de l'organisation des gènes sont décrites dans la **Figure 19.11**. Le transcrit commence par une séquence leader 5' suivie de la séquence du petit ARNr, puis vient une région appelée **intervalle transcrit** (spacer) et finalement la séquence du grand ARNr jusqu'à l'extrémité 3' de la molécule.

(Il ne faut pas confondre *intervalle transcrit* avec *intervalle non transcrit* qui sépare les unités de transcription. Son nom reflète le fait qu'il fait partie de l'unité de transcription, mais n'est pas représenté dans les produits matures de l'ARN.)

L'unité de transcription est donc plus longue que la longueur combinée des ARNr matures. Le **Tableau 19.3** résume quelques exemples de relations entre le transcrit primaire et les ARNr.

L'unité de transcription est plus courte chez les bactéries ; les séquences d'ARN ribosomal y constituent 80 % de la longueur totale qui est de 6 000 bases. On n'observe aucun schéma spécifique dans une batterie d'eucaryotes ; la longueur de transcription varie entre 7 000 et 8 000 bases, avec 70 à 80 % de la séquence représentant les ARNr. Le précurseur est plus long chez les mammifères (où il est appelé ARN 45S, d'après son coefficient de sédimentation). Dans ce cas, les séquences matures d'ARNr n'occupent que 50 % du transcrit entier.

Qu'arrive-t-il aux parties non ribosomales de l'ARN

Tableau 19.3

Chacun des précurseurs des ARNr est plus long que les deux ARNr matures.

Espèces	Longueur du transcrit (pb)	Longueur		% de l'ARNr dans le transcrit
		Grand	Petit	
<i>E. coli</i>	5 600	2 904	1 542	80 %
<i>S. cerevisiae</i>	7 200	3 750	2 000	80 %
<i>D. melanogaster</i>	7 750	4 100	2 000	78 %
<i>X. laevis</i>	7 875	4 475	1 925	79 %
<i>N. tabacum</i>	7 900	3 700	1 900	71 %
<i>G. domesticus</i>	11 250	4 625	1 800	57 %
<i>M. musculus</i>	13 400	4 712	1 950	52 %

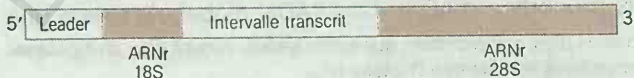
précurseur ? Les régions « leader » et l'intervalle transcrit sont éliminés pendant la maturation de l'ARNr. L'intervalle transcrit contient quelques séquences courtes qui sont libérées par clivage. Chez les mammifères et les amphibiens, une courte séquence forme l'ARN 5,8S, une petite molécule fixée par liaisons hydrogène à l'ARNr 28S dans le ribosome. Chez les bactéries, comme nous le verrons ultérieurement, des séquences d'ARNt peuvent se trouver dans l'intervalle transcrit et (parfois) également à l'extrémité 3' du transcrit primaire. Le reste de l'intervalle transcrit et toutes les séquences « leader » sont probablement dégradés en nucléotides.

Un groupe d'ADNr contient de nombreuses unités de transcription, chacune séparée de la suivante par un intervalle non transcrit. L'alternance d'une unité de transcription et d'un intervalle non transcrit est directement observée grâce à la microscopie électronique. L'exemple que nous montre la **Figure 19.12** est tiré d'un triton *N. viridescens*, dans lequel chaque unité de transcription est intensément exprimée ; de ce fait de nombreuses molécules d'ARN polymérase sont simultanément engagées dans la transcription d'une unité de répétition. Les polymérases sont si étroitement entassées les unes sur les autres que l'allongement de leurs produits forme une « matrice » caractéristique qui se déplace le long de l'unité de transcription.

La longueur de l'intervalle transcrit varie considérablement selon les espèces et (parfois) même à l'intérieur d'une même espèce. Le **Tableau 19.4** résume la situation.

Chez la levure il existe un court intervalle non transcrit dont la longueur est relativement constante. Chez *D. melanogaster*, la longueur de l'intervalle non transcrit varie presque du simple au double entre les différentes copies de l'unité de répétition. Une situation similaire s'observe chez *X. laevis*. Dans chacun de ces cas, toutes les unités de répétition sont présentes sous la forme d'un groupe en tandem unique situé sur un chromosome

Figure 19.11.
Les séquences des petites et des grandes molécules d'ARNr sont contenues dans une seule unité de transcription.



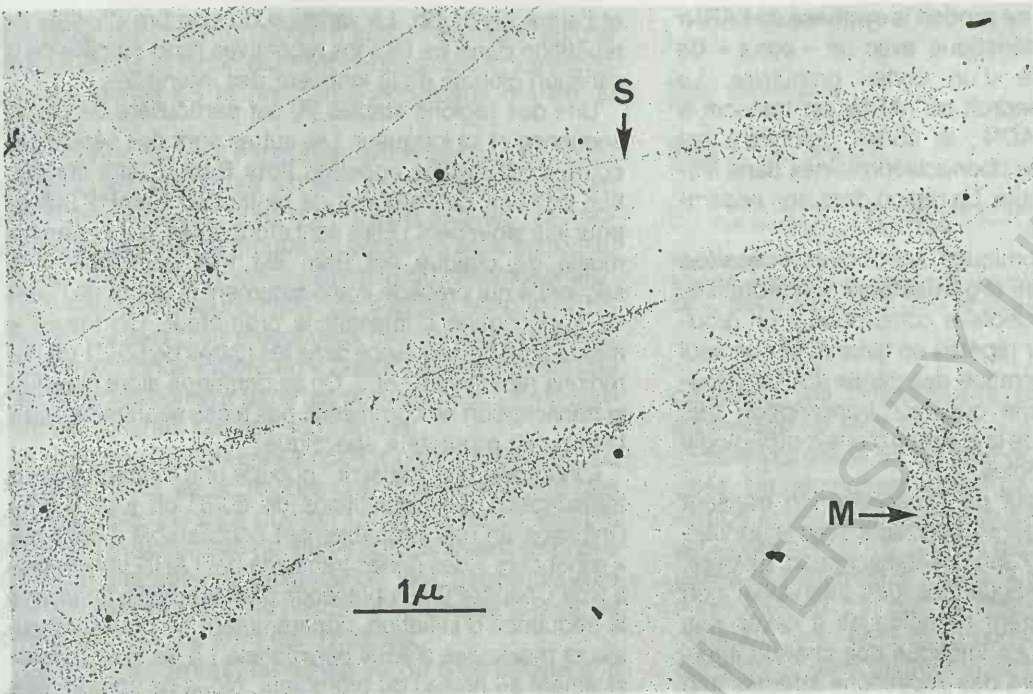


Figure 19.12
 La transcription des groupes d'ARNr engendre une série de matrices correspondant chacune à une unité de transcription séparée de la suivante par l'intervalle non transcrit.
 Photographie gracieusement fournie par Oscar Miller.

particulier. [Dans l'exemple de *D. melanogaster* celui-ci est le chromosome sexuel. Le groupe situé sur le chromosome X est plus grand que celui situé sur le chromosome Y ; les mouches femelles possèdent donc plus de copies des gènes d'ARNr que les mouches mâles (voir le Tableau 19.2). Cependant, tous les gènes ne sont pas actifs, comme nous le verrons ultérieurement.]
 Chez les mammifères l'unité de répétition est beaucoup plus grande ; elle comprend l'unité de transcription d'en-

viron 13 000 pb et un intervalle non transcrit de 30 000 pb. En général, les gènes se trouvent dans plusieurs groupes dispersés — dans le cas de l'homme et de la souris, ils résident respectivement sur cinq et six chromosomes. Une question intéressante serait de savoir comment les mécanismes correctifs qui fonctionnent probablement à l'intérieur d'un groupe unique pour assurer la stabilité de la séquence d'ARNr sont capables de fonctionner lorsqu'il y a plusieurs groupes.

Tableau 19.4
 La longueur de l'unité de répétition en tandem des groupes d'ADNr est variable.

Espèces	Longueur de l'unité de répétition	Longueur de l'intervalle non transcrit	Longueur du transcrit
<i>S. cerevisiae</i>	8 950 pb	1 750 pb	7 200 pb
<i>D. melanogaster</i>	11 500-14 200 pb	3 750-6 450 pb	7 750 pb
<i>X. laevis</i>	10 500-13 500 pb	2 300-5 300 pb	7 875 pb
<i>M. musculus</i>	44 000 pb	30 000 pb	13 400 pb

Ces données font référence à des gènes continus. L'échelle serait plus importante si les gènes discontinus 28S de *D. melanogaster* étaient inclus.

La région du noyau où se produit la synthèse de l'ARNr a une apparence caractéristique avec un « cœur » de nature fibrillaire entouré d'un cortex granulaire. Le « cœur » fibrillaire est l'endroit où l'ARNr est transcrit à partir de la matrice d'ADN ; le cortex granulaire est constitué de particules de ribonucléoprotéines dans lesquelles l'ARNr est assemblé. La région dans son ensemble est appelée **nucléole**.

Les régions chromosomiques spécifiques associées au nucléole sont appelées **organiseurs nucléolaires**. Chaque organisateur nucléolaire correspond à un groupement de gènes d'ARNr répétés en tandem sur un seul chromosome. La concentration des gènes d'ARNr répétés en tandem, de même que leur transcription très intense, est responsable de la création d'une morphologie caractéristique des nucléoles.

La variation de longueur de l'intervalle non transcrit dans un seul groupe de gènes contraste avec la conservation de la séquence de l'unité de transcription. En dépit de cette variation, les séquences des intervalles non transcrits plus longs restent homologues à celles des intervalles plus courts. Cela implique que chaque intervalle non transcrit contient des *répétitions internes* afin que la variation de longueur soit le résultat de modifications du nombre de répétitions d'une quelconque sous-unité.

La nature générale de l'intervalle non transcrit est illustrée par l'exemple de la **Figure 19.13**, qui montre son organisation chez *X. laevis*. Des régions dont les longueurs sont constantes alternent avec des régions dont les longueurs sont variables. Chacune des trois régions répétitives comprend un nombre variable de répétitions d'une séquence relativement courte. Une catégorie de ces régions possède les répétitions d'une séquence de 97 pb ; pour l'autre type situé en deux endroits, l'unité de répétition se trouve sous deux formes, l'une de 60 pb

et l'autre de 81 pb. La variation du nombre d'unités de répétition dans les régions répétitives rend compte de la variation globale de la longueur des intervalles.

Une des régions stables (A) est particulière de par sa séquence et sa longueur. Les autres sont des séquences courtes constantes appelées **îlots Bam**. (Cette appellation provient de l'enzyme de restriction BamH1 utilisée pour leur isolement.) Elles sont étonnantes, car la dernière moitié de chaque îlot Bam est très semblable à la séquence qui précède immédiatement le début de l'unité de transcription contenant le promoteur. On trouve le même type de séquence dans les unités de 60/81 pb des régions répétitives 2 et 3. On se demande alors pourquoi la transcription ne commence pas à ces séquences aussi bien qu'au promoteur spécifique.

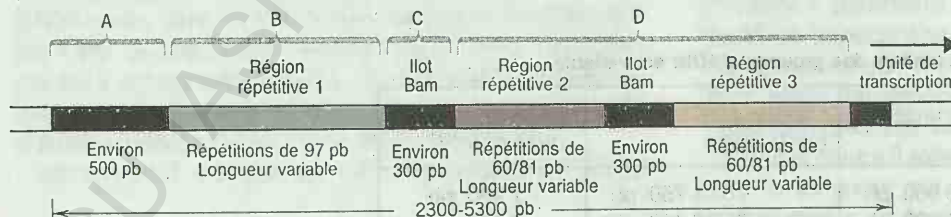
L'intervalle non transcrit * occupe une fonction dans la transcription. Les répétitions de 60/81 pb jouent dans l'initiation un rôle probablement comparable à celui des « enhancers » des gènes transcrits par l'ARN polymérase II (voir Chapitre 9). La présence d'un intervalle augmente la fréquence d'initiation, apparemment en poussant plusieurs molécules d'ARN polymérase I à se fixer successivement au niveau du promoteur. Nous ne comprenons pas encore la nature exacte de cette relation, ni de quelle manière elle dépend des homologies de séquence entre l'intervalle et le promoteur.

GÈNES 5S ET PSEUDOGÈNES SONT INTRIGUÉS

Chez la plupart des eucaryotes, l'organisation des gènes 5S se distingue de celle des gènes principaux d'ARNr. Le nombre total de gènes 5S peut être plus important, leur(s) localisation(s) différente(s), et ils sont transcrits par l'ARN polymérase III au lieu de l'ARN polymérase I

Figure 19.13

L'intervalle non transcrit de l'ARNr de *X. laevis* possède une structure riche en répétitions internes qui est responsable des variations de sa longueur.



* N. d. T. : Il semble que l'intervalle non transcrit contienne immédiatement en amont du promoteur du gène suivant, un terminateur de la transcription au niveau duquel s'arrête la molécule d'ARN polymérase ayant transcrit le gène précédent. Il est possible que pour des raisons de proximité, l'ARN polymérase libérée de l'ADN au niveau d'un signal de fin de transcription soit immédiatement disponible pour se fixer sur le promoteur voisin et initier la transcription d'un nouveau gène. Quoi qu'il en soit, ce couplage arrêt/réinitiation de la transcription semble important pour l'expression à un haut niveau des gènes ribosomiaux.

(l'enzyme nucléolaire). Nous n'avons vraiment qu'une petite idée de la façon dont est coordonnée la transcription des gènes 5S et des gènes principaux d'ARNr, si elle l'est.

Les gènes 5S sont habituellement organisés en un ou des groupes répétés en tandem. Ils ont été analysés en détail chez *X. laevis*, chez qui presque tous les chromosomes possèdent un groupe de gènes 5S situé au niveau ou très près du télomère (l'extrémité du chromosome).

On trouve plusieurs types de séquences d'ARN 5S chez le *X. laevis* ; ils forment deux catégories principales, dont la séquence diffère en 6 positions. On trouve l'une d'elles exclusivement dans les oocytes ; l'autre est synthétisée dans les cellules somatiques.

A peu près 20 000 gènes codent pour l'ARN 5S chez *X. laevis*, mais ils sont pour une grande majorité des représentants du type oocyte. La raison de ce nombre important pourrait être de fournir suffisamment de gènes 5S pour s'adapter à l'expression des gènes majeurs d'ARNr qui sont amplifiés pendant l'oogenèse. Les gènes d'ARN 5S, eux, ne sont pas amplifiés.

Le principal groupe de gènes 5S de type oocyte a la forme (familière) d'une alternance de gènes et d'intervalles non transcrits. Cependant l'unité de répétition comporte des répétitions internes, probablement à cause d'une duplication antérieure. Sa structure est montrée sur la **Figure 19.14**.

La séquence de l'intervalle est riche en A-T ; elle repose sur un nombre variable de répétitions d'une séquence de 15 pb et sur certains variants très apparentés. La longueur de chaque intervalle non transcrit varie donc selon le nombre de copies en tandem de cette répétition. La longueur moyenne de la séquence riche en A-T est approximativement de 400 pb.

Les 300 pb restantes de l'unité de répétition sont riches en G-C. Elles commencent par une séquence de 49 pb, suivie de la séquence de 120 pb du gène 5S (lui-même). Ensuite viennent à peu près une moitié et demie des unités de répétition riches en A-T, puis une autre copie de la séquence de 49 pb de l'intervalle riche en G-C. Puis il y a une répétition des premières 101 pb du gène, sous la forme d'une séquence incomplète qui n'est pas représentée dans l'ARN, et qui constitue un pseudogène. La région qui comprend le pseudogène représente une répétition de la séquence autour du gène actif, qui va de -73 à +101. Une caractéristique surprenante est que le groupe dans son ensemble doit contenir un nombre égal de gènes et pseudogènes.

Le promoteur de la transcription de l'ARN 5S se situe entièrement à l'intérieur du gène (voir Chapitre 9). La séquence du pseudogène diffère du gène actif en 9 points de mutation ; ces modifications sont probablement res-

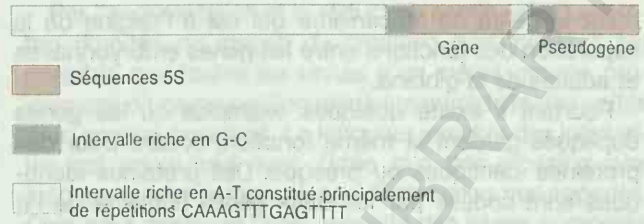


Figure 19.14

L'unité de répétition du principal groupe de gènes des oocytes 5S de *X. laevis* contient un gène et un pseudogène.

ponsables de l'impossibilité pour le pseudogène de synthétiser un ARN *in vivo*.

Tous les gènes 5S de type oocyte ne sont pas identiques. Un composant mineur de cet ARN, appelé ARN 5 « trace », révèle une certaine différence de séquence. Il est également codé par un déploiement en tandem de gènes et d'intervalles ; mais la longueur de la répétition n'est que de 350 pb, et l'intervalle n'a pas de point commun avec celui du groupe principal. Nous voyons donc qu'un gène ne peut subir que de légères modifications de séquence, tandis que l'intervalle change complètement. D'une même façon, un groupe de gènes 5S chez *X. borealis* possède des intervalles dont la séquence n'a aucun rapport avec celles mises en évidence chez *X. laevis*.

UN DILEMME DANS L'ÉVOLUTION : COMMENT LES COPIES MULTIPLES ACTIVES SE MAINTIENNENT-ELLES ?

On rencontre le même problème chaque fois qu'un gène a été dupliqué. Comment la sélection peut-elle s'imposer pour empêcher l'accumulation de mutations nuisibles.

La duplication d'un gène produit probablement un brusque relâchement de la pression de l'évolution sur sa séquence. Dès lors qu'il y a deux copies identiques, une modification dans la séquence d'une des deux ne privera pas l'organisme d'une protéine fonctionnelle, puisque la séquence d'origine de l'acide aminé continue d'être codée par l'autre copie. Ainsi la pression sélective sur les deux gènes est répartie et affaiblie jusqu'à ce que l'un des deux s'éloigne suffisamment de sa fonction originelle pour reporter toute la pression sélective sur l'autre.

Immédiatement après une duplication de gène, des modifications pourraient s'accumuler plus rapidement dans l'une des copies, conduisant éventuellement à une nouvelle fonction (ou à sa « mise à l'écart » sous la forme d'un pseudogène). Si une nouvelle fonction se développe, le gène peut alors évoluer au même rythme plus lent, caractéristique de la fonction originelle. C'est probable-

ment le genre de mécanisme qui est à l'origine de la séparation des fonctions entre les gènes embryonnaires et adultes de la globine.

Pourtant il existe quelques exemples où les gènes dupliqués gardent la même fonction, codant pour des protéines identiques ou presque. Des protéines identiques sont codées par les deux gènes humains de l' α globine, et il n'y a une différence que d'un seul acide aminé entre les deux protéines d' α globine. Cela n'est-il pas paradoxal alors que l'on considère qu'une duplication devrait être suivie par une variation ? Dans ces cas, nous pouvons probablement exclure une seule explication élémentaire qui veut que les deux gènes soient nécessaires et restent sous la pression sélective afin de produire suffisamment de protéine(s). Donc : comment la pression sélective est-elle maintenue ?

On a proposé deux mécanismes principaux. Ils ont en commun le principe selon lequel les gènes non alléliques ne sont pas transmis indépendamment mais qu'ils doivent tous deux être continuellement régénérés à partir d'une des copies de la génération précédente. Lorsqu'une mutation se produit dans une copie, soit elle est éliminée par hasard (parce que la séquence de l'autre copie prend le dessus), soit elle se propage dans les deux répliques (parce que la copie mutante devient la version dominante). La propagation expose une mutation à la sélection. Le résultat est que deux gènes évoluent ensemble comme s'il n'existait qu'un seul locus. On appelle cela **coïncidence** ou **évolution concertée** (parfois encore **coévolution**).

Un des mécanismes suppose que les séquences des gènes non alléliques sont directement comparées entre elles et homogénéisées par des enzymes qui reconnaissent toutes les différences. Cela peut être réalisé par l'échange de simples brins entre gènes, afin de former des gènes dont l'un des brins vient d'une copie et l'autre d'une autre. Toute différence se révèle par un appariement « impropre » des bases ; celui-ci peut retenir l'attention d'enzymes capables d'exciser et de remplacer une base, de sorte que seules les paires A-T et G-C survivent. Ce type d'événement est appelé **conversion génique** et est associé à la recombinaison génétique décrite au Chapitre 28.

Il existe un autre mécanisme qui veut que les deux copies du gène soient physiquement régénérées d'après une seule copie. Cela pourrait se produire par un **crossing-over** inégal. A la suite du type d'événement décrit à la Figure 19.3, par exemple, le chromosome portant un triple locus pourrait subir la délétion d'un des gènes. Des deux gènes restants, un et demi représente la séquence d'une des copies d'origine ; la moitié seulement de la séquence de l'autre copie originelle a survécu. Toute

mutation dans la première région est, désormais, présente dans les deux gènes et est sujette à la pression sélective. Ce type d'événement est appelé **fixation (ou implantation) d'un gène par crossing-over**.

Nous devrions pouvoir confirmer le schéma d'ensemble de ces mécanismes en comparant les séquences des gènes dupliqués. S'ils sont sujets à une évolution concertée, nous ne devrions pas voir parmi eux d'accumulation de substitutions des sites silencieux, car le processus d'homogénéisation s'exerce sur ces sites comme sur les sites de substitution. Nous savons que le rayon d'action du mécanisme de maintenance n'a pas besoin de s'étendre au-delà du gène lui-même, puisqu'il y a des cas de gènes dupliqués dont les séquences flanquantes sont entièrement différentes de chaque côté. En effet, on peut distinguer des frontières abruptes délimitant les extrémités des séquences ayant été homogénéisées.

Nous devons garder présent à l'esprit que l'existence de tels mécanismes peut invalider le sens de l'histoire de ces gènes par le biais de leur divergence, *parce que la divergence ne reflète que le temps passé depuis le dernier événement d'homogénéisation-régénération et non la duplication originelle*.

Lorsqu'il existe de nombreuses copies d'un gène, les effets immédiats d'une mutation sur n'importe quelle copie peuvent être très légers. Les conséquences d'une mutation individuelle sont diluées par le grand nombre de copies du gène qui maintiennent la séquence à l'état sauvage. Cela suggère qu'une quantité appréciable de copies mutantes pourraient s'accumuler avant que l'effet ne devienne assez important pour qu'elles soient éliminées par l'évolution.

L'élimination devient quantitative, une conclusion renforcée par le fait que la moitié des unités du groupe d'ADNr de *X. laevis* ou de *D. melanogaster* peut être supprimée sans effet pervers. De quelle manière ces unités se protègent-elles donc d'une accumulation progressive de mutations invisibles ? Et quelles sont les chances qu'a une mutation favorable (rare) de diffuser ses avantages dans le groupe ?

La conclusion inévitable est qu'il doit exister un mécanisme permettant à des variants spécifiques d'être minutieusement examinés par l'évolution. Deux types de mécanismes ont été proposés.

Le modèle de **correction instantanée** implique que, très souvent, chaque groupe entier de gènes soit remplacé par un nouveau jeu de copies qui vient d'une, ou tout au moins d'un très petit nombre de copies présentes à la génération précédente. Pour qu'une force sélective suffisante s'impose, « très souvent » pourrait n'être que « quelques générations » ; mais en termes pratiques, tout mécanisme doit être construit sur une base récurrente

régulière, c'est-à-dire que « très souvent » doit équivaloir à : « à chaque génération ».

Le petit nombre de copies qui donne naissance au nouveau groupe forme un ensemble « directeur » sur lequel agit la sélection ; les copies supprimées sont écartées de la sélection. Les copies « directrices » pourraient constituer un ensemble spécifique ou bien être choisies au hasard.

La difficulté avec ce modèle est qu'il prévoit la régénération du groupe à partir d'un *petit nombre d'unités de répétition*. Cela signifie que les unités de transcription et les intervalles doivent être régénérés, de sorte que l'unité de répétition entière devrait être de séquence homogène ou tout au moins devrait ne pas révéler plus de variations que celles correspondant au petit nombre de copies « directrices » à partir desquelles a été régénéré le groupe entier. Mais en fait, les intervalles présentent de continues modifications, alors que seules les unités de transcription sont constantes.

Le modèle de **fixation (ou implantation) d'un gène par crossing-over** suppose que le groupe entier soit sujet à un réarrangement perpétuel opéré par le mécanisme du crossing-over inégal. Cela s'applique essentiellement aux groupes en tandem, reproduisant sur une grande échelle les mécanismes intervenant de façon plus sporadique et sur une plus petite échelle au niveau des gènes de globine. Nous examinerons leurs conséquences de façon plus détaillée au Chapitre 21, mais il est, dès à présent, évident que les groupes en tandem fournissent de fréquentes occasions « d'erreurs d'appariement » de gènes dont les séquences sont effectivement les mêmes mais qui sont localisées dans des emplacements différents à l'intérieur de leurs groupes.

Par l'augmentation et la réduction continuelle du nombre d'unités au moyen du crossing-over inégal, il est possible que toutes les unités d'un même groupe proviennent d'une proportion relativement réduite des unités d'un groupe ancestral. La longueur variable des intervalles est en accord avec l'idée que des crossing-over inégaux surviennent au niveau d'intervalles mal appariés. Cela peut expliquer l'homogénéité des gènes comparée à la variabilité des intervalles. En effet les gènes sont exposés à la sélection, même lorsque les unités individuelles répétées sont amplifiées à l'intérieur du groupe ; mais les intervalles n'étant pas concernés peuvent accumuler les mutations.

POUR EN SAVOIR PLUS

Le système de la globine a été passé en revue par Maniatis et al. (*Ann. Rev. Genet.* 14, 145-178, 1980), et les bases de la thalassémie mises à jour par Weatherall & Clegs (*Cell* 29, 7-9, 1982). Les principes généraux de l'évolution des gènes et des protéines ont été récapitulés par Wilson et al. (*Ann. Rev. Biochem.* 46, 573-639, 1977). La notion de pseudogène a été revue par Little (*Cell* 28, 683-684, 1982). L'historique du décryptage de la structure des gènes d'histones, et des diverses estimations sur son degré de conservation, est retracé par une série de revues générales, de Hentschel & Birnstiel (*Cell*, 25 301-313, 1981) à Kedes (*Ann. Rev. Biochem.* 48, 837-870, 1979). La structure des groupes d'ADNr eucaryotes est traitée dans le Lewin, *Gene Expression*, 2, *Eucaryotic Genomes* (Wiley, New York, 1980, pp. 875-906). Reeder (*Cell* 38, 349-351, 1984) a analysé les intervalles présents dans l'ADN ribosomal.

CHAPITRE 20

DES GÉNOMES CONFINÉS DANS LES ORGANITES

On ne trouve les opérons que chez les procaryotes, mais dans certaines situations une région définie d'ADN eucaryote contient une batterie de gènes qui partagent la tâche de contribuer au même aspect du phénotype. Les mitochondries et les chloroplastes possèdent tous deux des génomes d'ADN qui codent pour toutes les espèces d'ARN et certaines protéines impliquées dans les fonctions de l'organite.

L'organisation du génome d'un organite peut être plus proche de la construction d'un génome viral que de celle d'un génome cellulaire, puisque ce génome ne code que pour certaines des fonctions (et non toutes) indispensables pour perpétuer l'organite. Les autres fonctions sont codées dans le noyau, exprimées par la voie du mécanisme synthétique des protéines cytoplasmiques, puis importées dans l'organite. Bien entendu, l'organite est un composant permanent de la cellule et non un virus qui ne ferait que passer. En fait, le génome de l'organite comprend une certaine longueur d'ADN qui, ayant été physiquement confinée dans une partie bien précise de la cellule, est donc soumise à son propre système d'expression et de régulation.

Le terme **d'hérédité cytoplasmique** a souvent été employé pour décrire le comportement des gènes dans les organites. Cependant, nous n'utiliserons pas cette définition, car il est important d'être capable de distinguer les phénomènes se déroulant dans le cytoplasme en général de ceux qui ont lieu spécifiquement dans les organites. Un terme général plus approprié pour évoquer

les gènes ne résidant pas à l'intérieur du noyau est **extranucléaire**, qui nous laisse la possibilité d'utiliser le terme de **synthèse des protéines cytoplasmiques** pour décrire le stade final de l'expression des gènes *nucléaires* ; tandis que les gènes de l'organite sont transcrits et traduits dans le *même* compartiment de l'organite, celui-là même où ils sont localisés.

LES GÈNES DE L'ORGANITE POSSÈDENT UNE TRANSMISSION NON MENDÉLIENNE

La première preuve de la présence de gènes à l'extérieur du noyau a été fournie par la transmission non mendélienne chez les plantes (que l'on observa dans les premières années de ce siècle, juste après la redécouverte de la transmission héréditaire de type mendélien).

La transmission non mendélienne se traduit par des proportions anormales de ségrégation lorsqu'un croisement est effectué entre le type mutant et le type sauvage. Sa forme extrême est la transmission uniparentale lorsque le génotype d'un seul parent est transmis héréditairement et que celui de l'autre est systématiquement perdu. Dans des exemples moins caricaturaux, la descendance de l'un des génotypes parentaux l'emporte sur l'autre.

Habituellement, le génotype de la mère est préférentiellement transmis. Cet effet est parfois décrit comme **l'hérédité maternelle (transmission maternelle)**. Un point important est que le génotype fourni par un parent

de sexe donné prédomine, ce qui contraste avec les comportements de la génétique mendélienne où des croisements réciproques témoignent de la participation des deux parents pour une transmission équilibrée (voir Chapitre 1).

La représentation de l'information génétique de l'organe dans un zygote dépend des contributions relatives de chaque parent ou des capacités différentielles de maintien d'une information (ou des deux). Cela contraste avec l'information génétique nucléaire qui provient équitablement de chaque parent.

Il est possible de prévoir l'hérédité maternelle chez les animaux supérieurs si l'on suppose que les mitochondries sont probablement entièrement transmises par l'ovule et pas du tout par le sperme. Ainsi, les gènes mitochondriaux proviennent exclusivement de la mère, et chez les mâles ils sont supprimés à chaque génération.

Il existe une cause différente d'hérédité uniparentale dans l'algue *Chlamydomonas reinhardtii*, dans laquelle des gamètes de tailles égales fusionnent pour former le zygote. Les deux parents transmettent l'ADN du chloroplaste dans un zygote. Cependant, l'ADN du chloroplaste fourni par un parent est dégradé peu après la formation du zygote, laissant à l'ADN de l'autre parent la possession exclusive de l'organe.

A cause de l'existence de ces deux mécanismes il est donc impossible de déterminer la raison de la prédominance d'un génotype parental dans chaque cas particulier (par exemple, chez les plantes supérieures). Même lorsque les parents transmettent des quantités différentes de cytoplasme au zygote, même si l'un des parents fournit probablement plus de copies du génome d'organe que l'autre, des mécanismes sélectifs pourraient intervenir pour altérer encore davantage l'équilibre.

Chez les plantes dont les deux génomes se maintiennent, les phénotypes « mutant » et « sauvage » subissent une ségrégation pendant la croissance somatique. Ainsi dans une seule plante hétérozygote, certains tissus peuvent avoir un seul phénotype parental tandis que d'autres présentent un phénotype de l'autre parent. Cette **ségrégation somatique** contraste avec la stabilité du génotype nucléaire entier, répondant aux lois de la génétique mendélienne.

L'association de la ségrégation somatique et de la transmission non mendélienne suggère qu'il existe de multiples copies de gènes d'organe dans le zygote, qui seront réparties dans différentes cellules pendant la division somatique.

Il n'est pas sans importance de pouvoir identifier l'organe véhiculant tel ou tel marqueur génétique. Puisque le seul organe, parmi les cellules des animaux

supérieurs à posséder de l'ADN est la mitochondrie, c'est probablement l'unique localisation extranucléaire du matériel génétique. Mais chez les plantes et chez certains eucaryotes unicellulaires, chloroplastes et mitochondries sont à la fois présents. Ce sont les seuls organites possédant de l'ADN ; il est probable (bien que non prouvé) que tout marqueur à la transmission non mendélienne soit localisé dans l'un ou l'autre de ces compartiments. Dans certains cas, la localisation apparaît clairement d'après la nature du phénotype mutant ; mais dans d'autres cas, l'altération est d'une forme plus courante rendant impossible l'identification du compartiment où se trouve le gène impliqué.

LES GÉNOMES DE L'ORGANITE SONT DES MOLÉCULES CIRCULAIRES D'ADN

La majorité des génomes d'organe ayant été caractérisés jusqu'à présent, sont sous la forme d'une seule molécule d'ADN de séquence particulière. Habituellement on peut extraire cette molécule sous la forme d'un double brin circulaire, mais des cassures pendant l'isolement surviennent parfois si fréquemment, que la majorité du matériel est récupéré sous forme de fragments linéaires. Les protozoaires ciliés fournissent une des exceptions où l'ADN mitochondrial est une molécule linéaire.

En principe, plusieurs copies du génome sont présentes dans l'organe individuel. Comme il y a une multitude d'organites par cellule, il peut y avoir un grand nombre de génomes d'organites dans chacune d'elles. Bien que le génome d'organe en lui-même soit particulier, il constitue une séquence répétitive comparée à toute séquence nucléaire non répétitive.

Les génomes du chloroplaste sont relativement grands, habituellement d'environ 140 kb chez les plantes supérieures, et de moins de 200 kb chez les eucaryotes inférieurs. Cela est comparable à la taille d'un grand bactériophage, par exemple, T4 qui est, approximativement, de 165 kb. Le **Tableau 20.1** montre que deux eucaryotes inférieurs comptent de nombreuses molécules d'ADN circulaires par chloroplaste. La quantité totale d'ADN dans les chloroplastes constitue plusieurs % de l'ADN cellulaire total. Malheureusement, nous n'avons pas d'exemples comparables pour les plantes supérieures, mais dans ce cas la quantité d'ADN nucléaire est beaucoup plus importante, et donc la proportion de la masse présente dans le chloroplaste doit être relativement petite.

La taille totale des génomes mitochondriaux varie d'un peu près un ordre de grandeur, comme nous pouvons le voir dans le **Tableau 20.2**. Les mammifères ont des

Tableau 20.1

L'ADN des chloroplastes montre une gamme étroite de complexités.

Espèces	Complexité d'ADN	Génomes par organe	Organites par cellule	ADN chloroplaste /ADN total
<i>C. reinhardtii</i>	195 kb	70-100	1	14 %
<i>E. gracilis</i>	135 kb	40	15	3 %
Plantes supérieures	140 kb	inconnu	inconnu	inconnu

génomés mitochondriaux réduits, assez petits pour avoir été complètement séquencés, chez l'homme, la souris et la vache ; tous mesurent environ 16,5 kb. Nous ne disposons de données sur le nombre d'organites par cellule, que pour les lignées de cellules en culture, mais ce nombre est élevé (plusieurs centaines). La quantité totale d'ADN mitochondrial est petite ; elle représente moins de 1 % de la quantité d'ADN nucléaire. On dénombre de 1 à 10 génomes par mitochondrie. La taille du génome mitochondrial de la mouche du vinaigre (*drosophile*) et de la grenouille est légèrement plus grande, mais nous manquons d'informations les concernant sur le nombre de génomes et d'organites.

Chez la levure, le génome mitochondrial est beaucoup plus grand. Chez *S. cerevisiae*, la taille réelle varie assez considérablement selon les différentes souches, mais est de quelque 84 kb. Il y a environ 22 mitochondries par cellule et approximativement 4 génomes par organe. Dans les cellules en cours de multiplication, la quantité d'ADN mitochondrial peut atteindre 18 % (Chez d'autres levures les génomes mitochondriaux peuvent être beaucoup plus grands.)

La taille de l'ADN mitochondrial des plantes est extrêmement variable, avec un minimum d'environ 100 kb. Nous n'avons aucune idée de la signification de cet ADN supplémentaire dans les plus grands génomes mitochondriaux. La taille du génome le rend difficile à isoler dans

son intégralité, mais la cartographie de restriction chez plusieurs plantes suggère que chaque génome possède une séquence unique, organisée en un cercle. A l'intérieur de ce cercle, il peut exister de courtes séquences homologues. Des recombinaisons entre ces éléments engendrent de plus petites molécules circulaires subgénomiques qui coexistent avec le génome entier, ce qui explique l'apparente complexité des ADN mitochondriaux de la plante.

LES ORGANITES EXPRIMENT LEURS PROPRES GÈNES

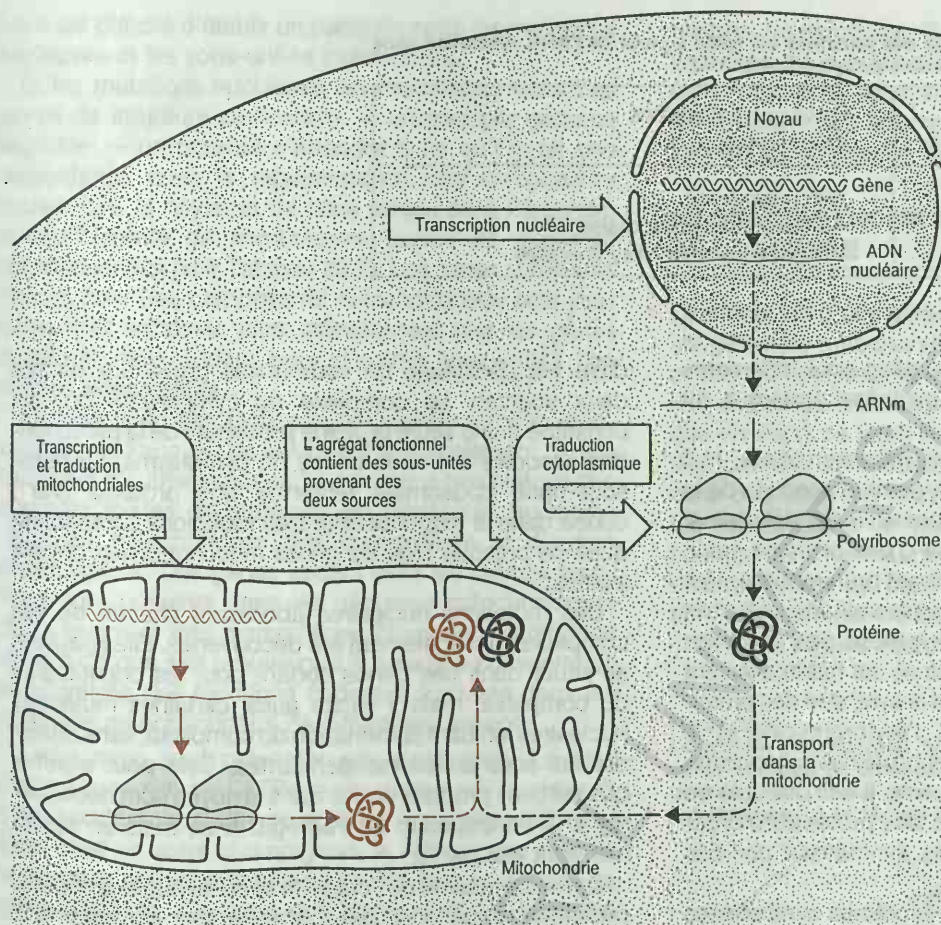
Les deux organites impliqués dans la conversion d'énergie suivent une économie mixte. La plupart des protéines les constituant sont importées du cytoplasme environnant, où leur synthèse représente le stade final d'expression des gènes nucléaires. Mais chaque organe s'implique également dans sa propre synthèse protéique. Comme le montre la **Figure 20.1** dans l'exemple de la mitochondrie, cet effort est centré sur la production d'un petit nombre de protéines, dont chacune est un composant d'un agrégat oligomérique comprenant aussi certaines sous-unités protéiques importées du cytoplasme.

La machinerie de synthèse protéique de l'organe est elle-même d'origine mixte. La plupart des protéines la

Tableau 20.2

La complexité de l'ADN mitochondrial est extrêmement variable.

Espèces	Complexité de l'ADN	Génomes /organe	Organites /cellule	ADN mit /ADN total
Souris (cellules L)	16,2 kb	2	500	0,2 %
Homme (Hela)	16,6 kb	10	800	1 %
<i>D. melanogaster</i>	18,4 kb	inconnu	inconnu	inconnu
<i>X. laevis</i>	18,4 kb	inconnu	inconnu	inconnu
<i>S. cerevisiae</i>	84 kb	4	22	18 %
<i>P. sativum</i>	110 kb	inconnu	inconnu	inconnu
Maïs	570 kb	inconnu	inconnu	inconnu

**Figure 20.1**

Les agrégats de protéines mitochondriales sont construits à partir à la fois des produits d'expression des gènes nucléaires et des gènes mitochondriaux.

composant sont transportées dans l'organite depuis le cytoplasme environnant. Cependant, les acides nucléiques ne peuvent traverser la membrane de l'organite, dans quelque direction que ce soit. Ainsi parmi les produits de l'organite lui-même, on peut compter tous les ARN faisant partie de la machinerie de synthèse des protéines.

L'origine de la machinerie des systèmes d'expression des gènes dans la mitochondrie de levure est résumée dans le **Tableau 20.3**. Ils sont généralement similaires dans les mitochondries des autres espèces et dans les chloroplastes. Dans la mitochondrie de levure, huit ARN ont été identifiés ; ils sont transcrits par une ARN polymérase, qui est (probablement) codée à l'intérieur du noyau. Seuls les ARN ayant été transcrits à partir des gènes

mitochondriaux sont traduits dans l'organite ; et vice-versa, c'est le seul compartiment cellulaire dans lequel ils peuvent être exprimés.

Le ribosome mitochondrial est constitué des habituelles petites et grandes sous-unités, chacune possédant un seul ARNr codé à l'intérieur de la mitochondrie, ainsi qu'un grand nombre de protéines importées du cytoplasme. Cependant, les structures du ribosome varient considérablement entre les espèces (voir le Chapitre 7). La batterie d'ARNt est plus petite qu'il n'est nécessaire dans le cytoplasme, comme nous le décrivons au Chapitre 6. Les aminoacyl ARNt synthétases responsables de la prise en charge des ARNt sont transférées du cytoplasme. Toutes les protéines de l'appareil de synthèse semblent être particulières à la mitochondrie ; bien que

Tableau 20.3
L'appareil mitochondrial de la synthèse des protéines est édifié à partir de l'ARN synthétisé dans la mitochondrie et des protéines importées du cytoplasme.

Constituant	ARN synthétisés dans la mitochondrie	Protéines synthétisées dans le cytoplasme
Matrice	~ 8 ARNm	ARN polymérase
Ribosome	2 ARNr	~ 75 protéines ribosomales
Adaptateur	22 ARNt	22 aminoacyl ARNt synthétases

codées par les gènes nucléaires, elles sont différentes des protéines de l'appareil de synthèse protéique du cytoplasme environnant.

Correspondant aux huit ARNm mitochondriaux, huit protéines peuvent être identifiées comme produits de la synthèse mitochondriale. Il est possible de déterminer « *in vivo* » l'origine de la synthèse d'une protéine grâce à sa sensibilité à des produits inhibant préférentiellement soit la synthèse des protéines cytoplasmiques, soit celle des protéines de l'organite. La synthèse des protéines du chloroplaste et des mitochondries est habituellement sensible aux antibiotiques inhibant la synthèse des protéines bactériennes, par exemple, l'érythromycine et le chloramphénicol. La synthèse des protéines cytoplasmiques est sensible à la cycloheximide. Il est possible de déterminer « *in vitro* » les origines des protéines d'organite en caractérisant les produits synthétisés par des préparations d'organites isolés.

Les complexes renfermant les protéines synthétisées dans la mitochondrie de levure sont résumés dans le **Tableau 20.4**. L'ATPase est constituée de deux unités, un facteur membranaire comportant 3 sous-unités codées par le génome mitochondrial, et l'ATPase F1 soluble constituée de 5 sous-unités synthétisées dans le cytoplasme. De la même manière, la cytochrome-c-oxy-

dase est constituée de sous-unités provenant de ces deux sources. Le complexe du cytochrome bc₁ est constitué d'une seule protéine provenant de la mitochondrie associée à 6 sous-unités du cytoplasme. La petite sous-unité ribosomale renferme une protéine (Var1) codée dans la mitochondrie. Des mutations caractéristiques de presque tous les gènes mitochondriaux ont été isolées.

Des mutations nucléaires abolissant chacun de ces complexes ont également été découvertes. Elles peuvent se situer dans des gènes codant pour les constituants du complexe, mais il existe aussi certaines mutations nucléaires inhibant la formation du complexe, sans atteindre les sous-unités indépendantes. Cela peut signifier que certains produits codés dans le noyau sont nécessaires à la construction de l'agrégat final, mais en réalité n'en font pas partie.

Dans les mutants de l'organite et du noyau, le complexe protéique ne parvient pas à s'assembler correctement, plusieurs sous-unités sont en conséquence absentes. Cela suggère que l'édification de chacun de ces complexes sous sa forme habituelle associée à la membrane est un processus compliqué dans lequel l'absence d'un des composants rend impossible l'assemblage des autres. La conséquence pratique de ces relations est qu'il

Tableau 20.4
Les complexes protéiques mitochondriaux (de la levure) sont édifiés à partir des protéines synthétisées dans différents compartiments cellulaires.

Complexe protéique	Masse (daltons)	Sous-unités fabriquées dans la mitochondrie	Sous-unités fabriquées dans le cytoplasme
ATPase sensible à l'oligomycine	340.000	ATPase 6,8,9 (facteur membranaire FO)	ATPase 1,2,3,4,7 (ATPase F1)
Cytochrome c. oxydase	137.000	CO 1,2,3	CO 4,5,6,7
Complexe du cytochrome bc 1	160.000	Apoprotéine du cytochrome b	6 sous-unités

nous est difficile d'établir un parallèle entre les mutations nucléaires et les sous-unités individuelles.

Si les mutations nucléaires sont localisées dans des gènes de structure particuliers — uniques par génome haploïde — on pourrait s'attendre à ce qu'il y ait une discordance dans la représentation des composants codés dans le noyau et de ceux codés dans l'organite, puisqu'il existe de nombreuses copies de génomes mitochondriaux (voir Tableau 20.2). Les gènes nucléaires codant pour des composants mitochondriaux sont probablement exprimés plus efficacement que les gènes mitochondriaux codant pour d'autres sous-unités des mêmes composants.

La répartition des tâches si particulière entre la mitochondrie et le noyau de levure est-elle également caractéristique d'autres espèces ? La composition globale des agrégats mixtes de protéines est semblable, au moins dans les mitochondries des mammifères et d'autres champignons. Des gènes codant pour les mêmes fonctions sont présents dans les génomes mitochondriaux de la levure et des mammifères. Nous nous attendons donc à ce que soit prouvé qu'il y a eu une conservation substantielle des fonctions codantes trouvées dans les génomes mitochondriaux de différentes espèces.

Cette règle peut présenter des exceptions. La plus petite sous-unité de l'ATPase est codée par le génome mitochondrial chez la levure et chez l'*Aspergillus nidulans*. Mais chez un autre champignon, le *Neurospora crassa*, la protéine correspondante se révèle être codée par un gène nucléaire. Ce gène doit donc avoir été transféré entre les génomes mitochondriaux et nucléaires.

Quels sont les gènes véhiculés par les chloroplastes ? Comme les mitochondries, ils codent pour toutes les espèces d'ARNt et d'ARNr nécessaires à la synthèse des protéines. Une cinquantaine de protéines sont synthétisées dans les chloroplastes ; des gènes ont été identifiés pour à peu près la moitié d'entre elles. Il semble que les gènes véhiculés par le génome des organites soient les mêmes, ou tout au moins soient très similaires dans les chloroplastes de la plupart des espèces. L'organisation des gènes en un génome circulaire est toutefois relativement différente dans plusieurs groupes de plantes et d'algues.

Les génomes des mitochondries et des chloroplastes sont-ils particuliers, ou bien certaines séquences sont-elles communes aux organites ou encore à un organite et au noyau ? La majeure partie d'un génome d'organite est propre à cet organite, mais une homologie de 12 kb a été découverte entre les ADN des mitochondries et des chloroplastes du maïs. Dans certains organismes, des séquences mitochondriales possèdent dans le noyau des

régions homologues. Des échanges d'ADN entre les organites ou avec le noyau sont sans nul doute rares, mais pourraient survenir de temps en temps. Nous ne connaissons pas encore les fonctions exercées par les séquences impliquées dans ces homologies.

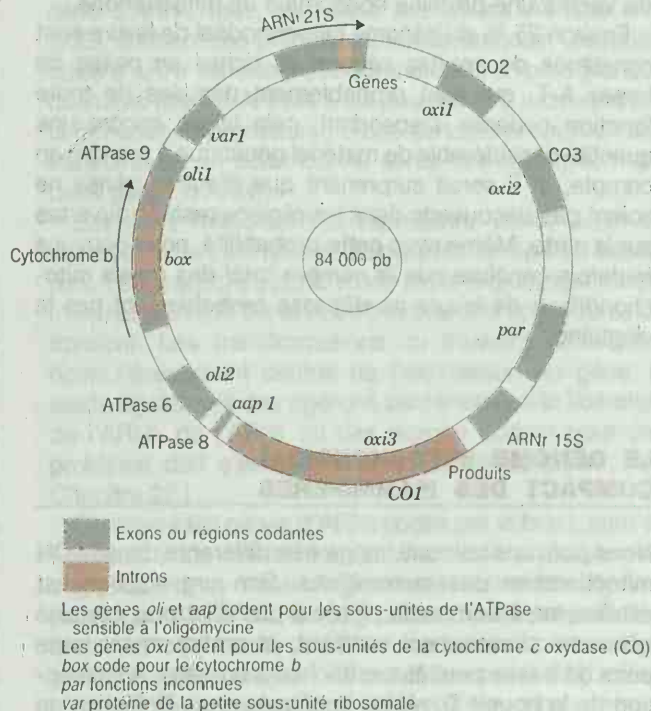
LE GRAND GÉNOME MITOCHONDRIAL DE LA LEVURE

Le simple fait d'une différence de taille du quintuple entre les génomes mitochondriaux de la *S. cerevisiae* (84 kb) et des mammifères (16 kb) nous indique qu'il doit exister une différence importante dans leur organisation génétique en dépit de leurs fonctions communes. Le nombre des produits synthétisés par voie endogène impliqués dans des fonctions enzymatiques mitochondriales semble être similaire. Le matériel génétique supplémentaire contenu dans les mitochondries de levure représente-t-il d'autres protéines, peut-être impliquées dans la régulation, ou bien reste-t-il inexprimé ?

La carte Figure 20.2 représente la majeure partie des

Figure 20.2

Le génome mitochondrial de la *S. cerevisiae* contient à la fois des gènes continus et discontinus codant pour des protéines, des gènes d'ARNr, et des gènes d'ARNt (dont les positions ne sont pas indiquées). Les flèches indiquent le sens de la transcription.



produits de l'ARN et des protéines de la mitochondrie de levure (excepté les quelque 22 molécules d'ARNt qui n'ont pas encore été complètement cartographiées). La caractéristique la plus significative est la dispersion des loci sur une carte.

Il est extrêmement rare que les gènes de l'ARNr soient séparés ; ce génome en possède la particularité. Le gène de l'ARNr 15S est interrompu et il est situé à environ 25 000 pb du gène de l'ARNr 21S. Dans certaines souches de *S. cerevisiae* ce dernier gène possède un intron unique (comme le montre la carte) ; dans d'autres souches, il est continu.

Les deux loci prédominants sont les gènes « mosaïque » *box* (codant pour le cytochrome *b*) et *oxi3* (codant pour la sous-unité 1 de la cytochrome oxydase). Ensemble, ces deux gènes sont presque aussi longs que le génome mitochondrial entier chez les mammifères ! Nombre des longs introns de ces gènes ont des cadres de lecture ouverts en phase avec l'exon précédent ; dans quelques cas au moins, il y a des arguments en faveur de la traduction de l'intron (voir Chapitre 23). Cela ajoute plusieurs protéines, probablement toutes synthétisées en petites quantités, au crédit de la mitochondrie de levure.

Les gènes restants semblent être ininterrompus. Ils correspondent aux deux autres sous-unités de la cytochrome oxydase codées par la mitochondrie, à la sous-unité (ou aux sous-unités) de l'ATPase, et (dans le cas de *var1*) à une protéine ribosomale de mitochondrie.

Environ 25 % du génome mitochondrial de levure sont constitués de courtes séquences riches en paires de bases A-T, qui sont probablement dénuées de toute fonction codante. Cependant, cela laisse encore une quantité considérable de matériel génétique à prendre en compte, et il serait surprenant que d'autres gènes ne soient pas découverts dans les régions restées ouvertes sur la carte. Même avec cette probabilité, nous pouvons toutefois conclure que le nombre total des gènes mitochondriaux de levure ne dépasse probablement pas la vingtaine.

LE GÉNOME MITOCHONDRIAL COMPACT DES MAMMIFÈRES

Nous pouvons voir une image très différente dans l'ADN mitochondrial des mammifères. Son organisation est extrêmement compacte ; il n'y a pas d'introns, certains gènes se chevauchent vraiment, et pratiquement toute paire de bases peut être rattachée à un gène. À l'exception de la boucle D, région impliquée dans l'initiation de

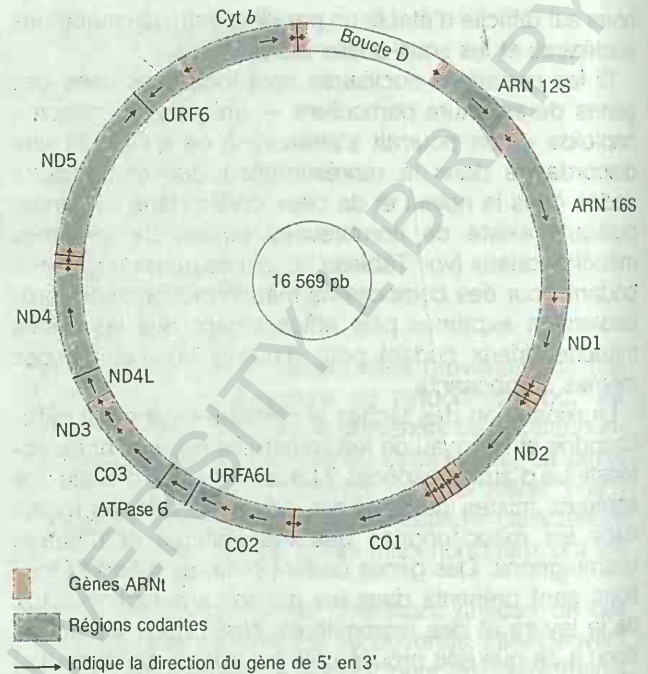


Figure 20.3

La séquence de l'ADN mitochondrial humain permet d'identifier 22 gènes d'ARNt, 2 gènes d'ARNr et 13 régions potentielles codant pour des protéines, dont certaines ont été identifiées.

CO indique la cytochrome oxydase, ND indique la NADH déshydrogénase. Des 15 régions codant pour des protéines ou pour l'ARNr, 14 sont transcrites dans le même sens (URF 6 en est l'exception). Des 22 gènes d'ARNt, 14 sont exprimés dans le sens des aiguilles d'une montre et 8 sont lus dans le sens contraire.

la réplication de l'ADN, pas plus de 87 des 16 569 pb du génome mitochondrial humain peuvent être considérées comme localisées dans des régions intercistroniques.

Les séquences nucléotidiques complètes des génomes mitochondriaux murins et humains révèlent une homologie considérable de leur organisation. La carte du génome humain est résumée sur la Figure 20.3. Il y a 13 régions potentielles codant pour des protéines. De même que la levure, elles incluent le cytochrome *b*, les trois sous-unités habituelles de la cytochrome oxydase, et une des sous-unités de l'ATPase. Contrairement à la levure, la mitochondrie de mammifère code pour six sous-unités (ou protéines associées) de NADH déshydrogénase. Les deux cadres de lecture qui restent sont indiqués URF (Unidentified reading frame = cadre de lecture non identifié) ; les fonctions de ces protéines (présomées) sont jusqu'à présent inconnues.

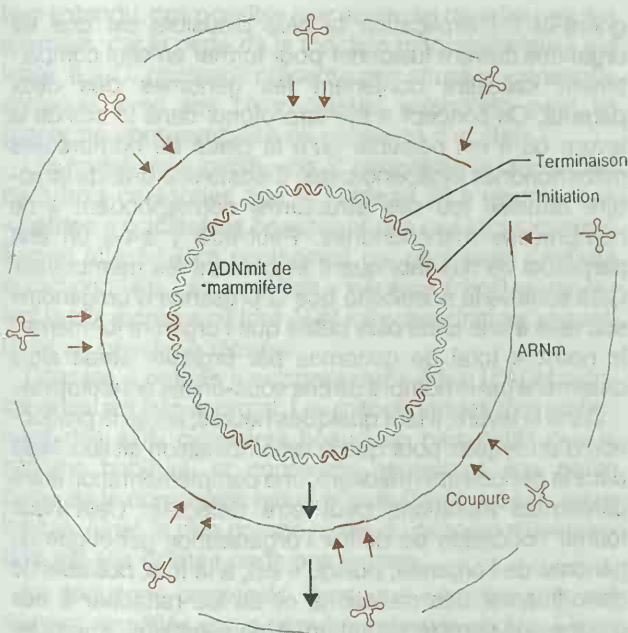


Figure 20.4
L'ADN mitochondrial humain est transcrit en un seul transcrit dont les ARNt sont coupés pour libérer les ARNr et les ARNm.

L'organisation des gènes montre des caractéristiques frappantes qui reflètent les particularités d'expression du gène dans la mitochondrie de mammifère. La plupart des gènes sont exprimés dans le même sens et les gènes d'ARNt sont localisés entre les gènes codant pour des ARNr ou des protéines.

Dans de nombreux cas, il n'y a aucune séparation entre les gènes. La dernière base d'un gène est adjacente à la première base du gène suivant. Dans certains cas, il y a chevauchement, plus couramment d'une seule base, de telle sorte que la dernière base d'un gène sert de première base au gène suivant. Cinq des cadres de lecture n'ont pas de codon de terminaison ; ces derniers finissent par U ou UA et, ainsi, un codon ocre est créé par la polyadénylation du transcrit (l'ARN messager mitochondrial acquiert une courte séquence poly(A) à son extrémité 3'). Dans trois cas, le codon de terminaison se révèle être AGA ou AGG, habituellement utilisé comme arginine dans le code génétique (voir Chapitre 5).

Les espèces d'ARNm, correspondant à toutes les régions codant pour des protéines, ont été identifiées, et dans tous les cas, un codon d'initiation se trouve, dans les six premières bases. Le codon utilisé pour l'initiation

peut apparemment être AUG, AUA ou AUU. Des régions 5' et 3' non traduites sont pratiquement absentes des ARNm mitochondriaux humains.

Presque tous les gènes sont exprimés dans le sens des aiguilles d'une montre, comme l'indiquent les flèches sur la carte. Dans un cas seulement on trouve deux régions codant dans le sens des aiguilles d'une montre sous la forme de gènes contigus (ATPase 6 et CO3). Dans tous les autres cas, au moins un gène d'ARNt sépare des régions codantes adjacentes.

La ponctuation par les gènes d'ARNt des régions codant pour les ARNr et les protéines ne laisse aucune place à des promoteurs comparables à ceux que l'on trouve dans les noyaux des eucaryotes ou les bactéries. Un seul promoteur pour la transcription dans le sens des aiguilles d'une montre se trouve dans la région de la boucle D. Cela nous suggère le modèle illustré sur la **Figure 20.4**. La transcription débute juste avant le gène d'ARNt situé devant le gène d'ARNr 12S, et se poursuit presque tout autour du cercle, pour se terminer dans la boucle D. Le brin transcrit est appelé **brin H** (pour heavy) [car l'ADN mitochondrial peut être séparé en simples brins lourds (H) ou légers (L) d'après leur densité].

La signification de l'alternance des gènes d'ARNt avec des gènes d'ARNr et des gènes codant pour les protéines est que les ARNt indiquent des sites de clivage. Si l'on coupe le transcrit primaire de part et d'autre de chaque gène d'ARNt, tous, sauf deux des gènes (ATPase 6 et CO3), donnent des produits monocistroniques. Les molécules d'ARNr semblent être synthétisées en plus grandes quantités que les ARNm. Cela pourrait être provoqué par la terminaison prématurée d'une certaine proportion des transcrits en un point donné après la transcription des deux gènes d'ARNr.

Cet ADN mitochondrial représente donc, pour les eucaryotes, la plus forte analogie avec un opéron bactérien. Il est transcrit à partir d'une seule région promotrice, mais les ARNt, ARNr et ARNm individuels sont libérés du transcrit. Les transformations du transcrit deviennent donc l'événement central de l'expression du gène. (Il existe également des opérons bactériens où la libération de l'ARNt, de l'ARNr ou des régions codant pour des protéines doit s'accomplir après la transcription ; voir Chapitre 22.)

Comment les gènes d'ARNt codés par le brin L sont-ils exprimés ? Un mécanisme similaire doit probablement s'appliquer ; un transcrit géant est vraisemblablement produit pour le brin L, les séquences d'ARNt, et peut-être certaines séquences d'ARNm, en seront conservées, pendant que les séquences restantes seront dégradées. L'origine de la transcription du brin L (en sens contraire

des aiguilles d'une montre) se situe également dans la région de la boucle D. On peut isoler de très grands transcrits du brin L ; leur maturation a donc probablement lieu un certain temps après la transcription. Nous ne trouvons pas de grands transcrits du brin H, ce qui suggère qu'ils subissent une maturation plus ou moins immédiatement après la transcription de chaque jonction d'ARNt.

UNE RECOMBINAISON A LIEU DANS (CERTAINS) ADN D'ORGANITE

La condition préalable à une recombinaison est la coexistence des génomes des deux parents, une condition rendue impossible dans les situations où l'absolue transmission uniparentale prédomine. Mais l'existence de cas où les deux génomes d'organite se maintiennent soulève immédiatement le problème de leurs éventuelles interactions. La complémentation est-elle possible entre mutants, une recombinaison peut-elle avoir lieu entre les génomes ?

La complémentation requiert l'enchevêtrement des produits d'expression du gène ; la recombinaison requiert une interaction physique entre les deux génomes. Ces deux exigences sont en quelque sorte en désaccord avec la vision statique de l'organite individuel, implicite si l'on décrit l'expression de ses gènes endogènes comme un phénomène strictement limité à son propre compartiment cellulaire. Pourtant, dans une espèce au moins, des recombinaisons peuvent avoir lieu entre des ADN de chloroplastes ; dans une autre, complémentation et recombinaison peuvent à la fois avoir lieu avec des génomes mitochondriaux.

Dans l'algue *C. reinhardtii*, l'ADN du chloroplaste est habituellement transmis d'un seul parent. Pourtant, dans la situation exceptionnelle où les gènes du parent habituellement prédominant ont été irradiés par rayonnement ultraviolet, les gènes du chloroplaste de l'autre parent survivent. Dans les zygotes biparentaux qui en résultent, la recombinaison a lieu entre les génomes des chloroplastes, et cela peut être utilisé pour établir une carte génétique.

Dans la levure *S. cerevisiae*, la recombinaison entre les marqueurs mitochondriaux est la norme. La nature de la recombinaison est inconnue, mais un croisement entre deux parents, dont les ADN mitochondriaux sont très différents, engendre une descendance possédant des ADN physiquement recombinants. Cela démontre directement que l'ADN mitochondrial d'un parent vient au contact de l'ADN mitochondrial fourni par l'autre parent.

Comment les ADN des organites parentaux se rejoignent-ils ? L'explication la plus plausible est que les organites doivent fusionner pour former un seul compartiment cellulaire contenant les génomes des deux parents. Ce concept a été approfondi dans le cas de la levure où il est possible qu'à la place de nombreuses mitochondries indépendantes, il existe une grande structure ramifiée (ou des structures) correspondant à de nombreuses mitochondries. Peut-être y a-t-il un état perpétuel de flux fabriquant et cassant les membranes. Cela soulève la possibilité que la présence d'un génome soit relié à une unité plus petite que l'organite lui-même ; le nombre total de génomes par organite serait alors déterminé par le nombre de ces sous-unités constitutives.

Dans la levure, il faut quelques heures, après la production d'un zygote, pour que la recombinaison ait lieu. Mais dans la période intermédiaire, une complémentation entre différentes mutations peut être détectée. Cela nous fournit l'occasion de définir l'organisation génétique du génome de l'organite, puisqu'il est, à la fois, possible de cartographier des mutations et de les rattacher à des groupes de complémentation. (Une semblable approche a également été couronnée de succès avec le chloroplaste du *C. reinhardtii*.)

LES RÉARRANGEMENTS DE L'ADN MITOCHONDRIAL DE LEVURE

Le génome mitochondrial de la levure a une structure d'une surprenante fluidité. Nous avons déjà signalé que l'on trouve certains gènes de formes différentes dans différentes souches de *S. cerevisiae*. En effet, la taille globale du génome mitochondrial peut varier substantiellement, jusqu'à 10 000 pb selon les souches. Des délétions étendues d'ADN mitochondrial peuvent se produire, comme celles que l'on trouve dans la variété « petite » de *S. cerevisiae* (et dans certaines autres levures). Elles peuvent apparaître spontanément, ou bien être induites par certains traitements, tels que l'addition de bromure d'éthidium.

Toutes les mutations « petite » sont dénuées de fonction mitochondriale. Leur existence est rendue possible par les modes de vie alternatifs de la levure qui leur permettent de survivre en aérobie (lorsque la respiration est essentielle), ou en anaérobie (lorsque l'on peut s'en dispenser). Ainsi, les mutations mitochondriales sont potentiellement léthales dans les cas où la croissance en aérobie est une condition obligatoire, alors que si la croissance en anaérobie est une condition facultative, l'organisme peut survivre.

La perte de la fonction mitochondriale oblige la levure à adopter le mode de vie anaérobie. Un tel choix n'est,

bien entendu, pas possible (par exemple) dans les cellules animales, car la perte de la fonction mitochondriale y est létale. (La *C. reinhardtii* présente une situation génétique analogue avec ses chloroplastes puisqu'elle peut se passer de photosynthèse en présence d'acétate.)

Les mutations « petite » de *S. cerevisiae* se regroupent dans trois catégories que le **Tableau 20.5** résume. Les « petites » **nucléaires** possèdent des mutations mendéliennes (c'est-à-dire nucléaires) qui abolissent l'activité mitochondriale. Les « petites » **neutres** (*rho*⁰) représentent le cas extrême où tout ADN mitochondrial est absent. C'est un génotype récessif.

Dans les « petites » **suppressives** (*rho*⁻) l'ADN mitochondrial est très anormal. Les mitochondries *rho*⁻ renferment des ADN circulaires beaucoup plus petits que le génome habituel, et contenant seulement une petite partie de la complexité habituelle de 0,2 % à 36 % selon les cas. Ainsi, la plus grande partie du génome mitochondrial est-elle déletée dans ces formes.

La séquence conservée dans une « petite » *rho*⁻ est souvent amplifiée en un grand nombre de copies. L'amplification se produit soit par accroissement de la ploïdie (du nombre de molécules d'ADN), soit par formation de molécules d'ADN multimériques, contenant chacune de nombreuses copies de la séquence. Les copies multiples sont agencées en tandem de répétitions directes ou inversées. La séquence de la « petite » ne représente pas nécessairement une région du génome mitochondrial, jadis adjacente, mais peut être composée de régions différentes juxtaposées. Certaines « petites » sont stables ; d'autres sont instables, et des réarrangements supplémentaires de la séquence peuvent se produire.

L'ADN conservé par une souche de « petite » peut se recombiner à l'ADN d'une autre « petite » ou d'une souche sauvage (appelée « grande » par comparaison). Cela permet aux marqueurs génétiques présents dans chaque « petite » d'être mis en relation avec la séquence d'ADN conservée. En fait, c'est l'utilisation de telles « petites » pour cette cartographie de délétion qui permet la réalisation de la carte génétique mitochondriale que montre la Figure 20.2.

N'importe quelle séquence d'ADN mitochondrial peut être conservée dans une « petite ». Cela implique que tous les segments de l'ADN peuvent être répliqués indépendamment. Nous ne savons pas comment ceci est possible, et si même cela dépend d'un signal de même nature que l'origine de réplication (voir Chapitre 13).

Lorsqu'une souche de « petite » *rho*⁻ de levure est « croisée » avec une souche sauvage, on ne trouve plus, dans une certaine proportion de la descendance, que le

Tableau 20.5
Les mutations « petite » font disparaître ou réarrangent l'ADN mitochondrial de la levure.

Catégorie	Nom	Type d'ADN mit
type sauvage	<i>grande</i>	~ 84 kb, circulaire
« petite » nucléaire	—	variable
« petite » suppressive	<i>rho</i> ⁻	ADN anormal
« petite » neutre	<i>rho</i> ⁰	pas d'ADN

génotype de la « petite ». L'ADN mitochondrial « sauvage » a disparu. Cet effet est appelé **suppression** ; et la proportion caractéristique de la descendance dans laquelle il a lieu détermine le **taux de suppression**.

La cause de la suppression dans les « petites » hautement suppressives est la réplication préférentielle de l'ADN mitochondrial de la « petite ». Cela se produit probablement parce que l'ADN de la « petite » possède une concentration accrue de séquences ressemblant aux origines de réplication, appelées régions *rep*. Nous voyons donc que l'ADN mitochondrial de la levure contient une séquence spécifique (ou plusieurs) qui peut être utilisée pour initier la réplication. Mais contrairement à son équivalent dans d'autres génomes, on peut se dispenser de cette séquence, puisque l'ADN d'une « petite » peut survivre sans elle (bien qu'elle soit alors moins suppressive). La cause de la suppression n'est donc pas nécessairement la même chez des « petites » moins suppressives.

Le comportement de l'ADN de « petite » et de l'ADN mitochondrial sauvage montre que l'organisation du génome présente une importante souplesse, à la fois quant à son expression et à sa réplication. Peut-être la différence avec d'autres ADN mitochondriaux vient-elle de la capacité de la levure à se dispenser totalement des fonctions mitochondriales. Il sera intéressant de caractériser les systèmes d'enzymes agissant sur cet ADN.

POUR EN SAVOIR PLUS

Gillham dans *Organelle Heredity* (Raven Press, 1978) a fourni une revue générale sur l'hérédité du chloroplaste et des mitochondries. Les points principaux de ce chapitre ont été passés en revue dans *Lewin's Gene Expression*, 2, *Eucaryotic chromosomes* (Wiley, New York, 1980, pp. 583-637). Clayton (*Ann. Rev. Biochem.* 53, 573-594, 1984) a résumé l'expression du génome mitochondrial des mammifères. Les progrès concernant le génome des chloroplastes ont été récapitulés par Palmer (*Ann. Rev. Genet.* 19, 325-354, 1985).

CHAPITRE 21

ORGANISATION DE L'ADN DE SÉQUENCE SIMPLE

Nombreuses sont les séquences du génome des eucaryotes ne possédant aucune fonction codante. Représentant pratiquement la majorité de l'ADN, ces séquences ont subi des contraintes évolutives différentes de celles imposées aux séquences codantes. Quelques-unes, bien sûr, font partie d'unités de transcription (par exemple, les régions bordantes non traduites de l'ARN messager ou les introns excisés durant la maturation de l'ARN messager).

Certaines autres peuvent fournir des signaux reconnus par des protéines comprenant des éléments tels des promoteurs transcriptionnels, des origines de réplication de l'ADN, des signaux de compaction du chromosome, des points d'ancrage du kinétochore et bien d'autres fonctions cellulaires. Bien peu de ces signaux ont été caractérisés jusqu'à pouvoir attribuer une fonction particulière à une séquence donnée. En fait, les promoteurs et les « enhancers » de transcription sont les seules séquences ayant été décrites. Et si l'on inclut les génomes viraux, on connaît également les origines de réplication de l'ADN. Le peu d'informations disponibles suggère que les signaux de contrôle de ce genre sont vraisemblablement assez courts.

Examinant par son autre aspect le problème de l'ADN non codant, nous pouvons nous demander ce que nous savons des courtes séquences qui sont répétées dans

le génome sous la forme de multiples copies identiques ou similaires ? Bien entendu cela renvoie à la définition de l'ADN répétitif tel qu'il a été caractérisé par les cinétiques de réassociation décrites au Chapitre 17. Pour récapituler, il existe deux sortes d'ADN répétitif :

- La fraction *moyennement répétitive* est constituée par une grande variété de familles de séquences dont les degrés de répétition et l'homogénéité varient d'une famille à l'autre. Les membres de ces familles sont souvent dispersés plus ou moins régulièrement au milieu de grandes régions d'ADN non répétitif. Une grande partie des composants moyennement répétitifs du génome des mammifères se révèle constituée des membres d'une seule famille discutée au Chapitre 30.
- A l'opposé, la fraction *hautement répétitive* est généralement constituée par de très courtes séquences répétées sous la forme de grands groupes d'éléments en tandem. Du fait de sa petite unité de répétition ceci est parfois décrit comme l'**ADN de séquence simple**. Ce type de composant est présent dans le génome de, pratiquement, tous les eucaryotes supérieurs mais dans des proportions extrêmement variables. Les données dont nous disposons sur l'évolution de ces composants nous renseignent sur les mécanismes impliqués dans les remaniements des séquences d'ADN au cours de longues périodes de temps, mais

jusqu'à présent nous n'avons pas de notions claires sur leur fonction, si elles en ont une.

L'ADN HAUTEMENT RÉPÉTITIF FORME DES SATELLITES

L'ADN hautement répétitif prend la forme de très courtes séquences, répétées de nombreuses fois en tandem. Un très long segment constitué de multiples copies de séquences hautement répétées forme un bloc de matériel. Dans certains cas, ces copies multiples sont identiques, dans d'autres, elles ne sont qu'apparentées.

La répétition en tandem d'une courte séquence crée souvent une fraction ayant des propriétés physiques propres qui permettent de l'isoler. Dans certains cas, la séquence répétitive possède une composition nucléotidique distincte de la composition moyenne du génome ; elle pourra alors être séparée grâce à sa densité à l'équilibre.

La densité à l'équilibre d'un double-brin d'ADN dépend de sa richesse en G-C selon la formule empirique.

$$\rho = 1,660 + 0,00098 (\% \text{ en G-C}) \text{ g-cm}^{-3} \quad (12)$$

La densité à l'équilibre est en général obtenue par centrifugation à travers un **gradient de densité** de CsCl. L'ADN forme une bande à la position correspondant à sa propre densité.

Lorsque l'on centrifuge, en gradient de densité, de l'ADN d'eucaryotes, la majeure partie du génome forme un continuum de fragments qui se regroupent en densité d'équilibre correspondant à la concentration moyenne en G-C du génome. C'est le **pic principal**. Il est quelquefois possible de détecter un (ou plusieurs) pic supplémentaire plus petit à des densités variables. Il s'agit de l'**ADN satellite**. Le gradient d'équilibre permet de séparer des fragments d'ADN différant de plus ou moins $0,005 \text{ g-cm}^{-3}$ ce qui correspond à une différence de contenu en G-C, d'environ 5 %.

L'ADN de souris fournit un bon exemple (**Figure 21.1**). Le graphique représente une analyse quantitative des bandes formées par centrifugation de l'ADN de souris à travers un gradient de densité de CsCl. La bande principale contient 92 % du matériel génomique, elle est en équilibre à une densité de $1,701 \text{ g-cm}^{-3}$ (ce qui correspond à un contenu en G-C de 42 %, typique des mammifères). Le petit pic représente 8 % du matériel génomique, sa densité propre est de $1,690 \text{ g-cm}^{-3}$. Il contient l'ADN satellite de la souris, dont le contenu en G-C (30 %) est beaucoup plus faible que celui de n'importe quelle autre partie du génome.

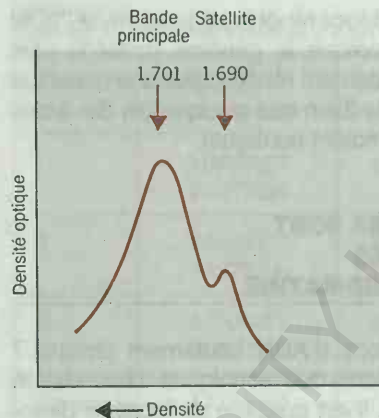


Figure 21.1
L'ADN de souris se sépare en gradient de densité en CsCl en une bande principale et un satellite.

Les satellites sont présents dans le génome de nombreux eucaryotes. Ils sont soit plus lourds, soit plus légers que la bande principale mais ne représentent que rarement plus de 5 % de l'ADN total. La centrifugation en gradient de Cs_2O_4 en présence d'ions Ag ou de différents réactifs, dont certains colorants, permet d'améliorer la résolution de l'ADN satellite. Celui-ci se résoud parfois en une seule bande distincte de la bande principale, parfois en de multiples séquences.

Le comportement de l'ADN en gradient de densité est le plus souvent anormal. En effet, lorsque l'on établit la composition nucléotidique d'un satellite, elle est le plus souvent différente de ce que sa densité d'équilibre ne le laissait prévoir. La raison en est que ρ est fonction non seulement de la composition en bases, mais aussi de sa structure, en termes de plus proches paires voisines. Il est probable que pour des séquences d'ADN simple, les conditions de relations aléatoires entre les paires, nécessaires au respect de l'équation 12, ne sont pas réunies. De plus, l'ADN satellite peut être méthylé, ce qui modifie sa densité.

La presque totalité de l'ADN hautement répétitif d'un génome peut être isolée sous forme de satellites. Lorsqu'un composant de cet ADN ne se sépare pas sous forme de satellite, il apparaît souvent, après l'avoir isolé, que ses propriétés sont semblables à celle de l'ADN satellite, c'est-à-dire qu'il est constitué de multiples unités de répétitions en tandem présentant des anomalies au cours de la centrifugation. Le matériel ainsi isolé est quelquefois dénommé **satellite cryptique**. L'ensemble des satellites, cryptiques ou apparents, constitue en

pratique la totalité des blocs répétés en tandem de l'ADN hautement répétitif. Lorsque le génome possède plus d'une sorte d'ADN hautement répété, chacune constitue son propre bloc satellite (bien que quelquefois des blocs de différentes natures soient contigus).

LES ADN SATELLITES SONT SOUVENT LOCALISÉS DANS L'HÉTÉROCHROMATINE

Où sont situés les blocs d'ADN hautement répétitif ? Grâce aux développements des techniques d'hybridation des acides nucléiques, il est possible de localiser directement les séquences satellites sur les chromosomes. Dans la technique d'**hybridation cytologique** ou **hybridation *in situ***, l'ADN chromosomique est dénaturé dans les cellules même après qu'elles ont été étalées sur des lames. On ajoute alors une solution contenant la sonde d'ADN ou d'ARN préalablement marquée par radioactivité. La sonde hybride avec ses compléments présents sur le génome dénaturé. Une autoradiographie permet de déterminer la position des points d'hybridation.

Les ADN satellites marqués sont souvent regroupés dans l'hétérochromatine présente autour des centromères des chromosomes en mitose. Le terme d'**hétérochromatine** désigne des régions des chromosomes en permanence étroitement enroulées et inertes, par opposition à l'**euchromatine** qui représente la majeure partie du génome (voir Chapitre 25).

La **Figure 21.2** illustre la localisation de l'ADN satellite sur les chromosomes de souris. Ici, les extrémités des chromosomes sont marquées, car c'est là que sont situés les centromères des chromosomes de *M. musculus*.

Cette localisation centromérique de l'ADN satellite suggère une possible fonction structurale dans le chromosome. Les centromères étant précisément les régions où sont formés, lors de la mitose et de la méiose, les kinétochores qui contrôlent les mouvements des chromosomes, cette fonction pourrait être en rapport avec le processus de ségrégation chromosomique.

C'est tout ce que l'on peut dire sur son rôle, qui, en dehors de cette hypothèse générale, demeure bien mystérieux. Conformément à leur séquence élémentaire et à leur structure condensée, les ADN satellites ne sont ni transcrits, ni traduits.



Figure 21.2

L'hybridation *in situ* montre la localisation centromérique de l'ADN satellite de souris.

Photographie gracieusement fournie par Mary Lou Pardue et Joe Gall.

LES SATELLITES DES ARTHROPODES
POSSÈDENT DE TRÈS COURTES UNITÉS
DE RÉPÉTITION IDENTIQUES

Chez les arthropodes (représentés notamment par les insectes et les crabes) chaque ADN satellite apparaît assez homogène. Habituellement, plus de 90 % du satellite sont composés par une seule et très courte unité de répétition. Il sera donc relativement facile d'en déterminer la séquence.

La « *Drosophila virilis* » possède trois satellites principaux et un satellite cryptique qui représentent ensemble une grande partie du génome (environ 40 %). Les séquences des satellites sont mentionnées dans le **Tableau 21.1**. Les séquences des trois satellites principaux sont très proches. A partir de la séquence du satellite I, la substitution d'une seule base suffit à produire les satellites II ou III. Le satellite cryptique dérive du satellite II par changement de 2 bases.

Le fait d'avoir trouvé la séquence satellite I dans d'autres espèces de *Drosophila* voisines de la « *Virilis* » pourrait indiquer que son apparition a précédé la séparation de ces espèces. Les séquences des satellites II et III semblent spécifiques de la « *D. virilis* » ; elles pourraient donc avoir évolué à partir du satellite I après séparation de cette espèce.

La caractéristique principale de ces satellites est la très petite taille de leur unité de répétition : 7 pb seulement. Des satellites de ce type sont présents dans d'autres espèces. Par exemple, la « *D. melanogaster* » qui possède un grand nombre de satellites dont plusieurs sont des unités de répétition de très petite taille (5, 7, 10 ou 12 pb). De semblables satellites existent également chez les crabes.

Une telle ressemblance entre les satellites d'une espèce telle qu'elle est notée chez « *D. virilis* » n'est pas une caractéristique générale. Dans d'autres génomes, les satellites peuvent avoir des séquences totalement différentes. *Chaque satellite s'est formé par amplification latérale d'une très courte séquence*. Cette séquence peut, soit dériver d'une séquence satellite préexistante (comme chez « *D. virilis* »), soit avoir une toute autre origine.

Il semble donc probable que les satellites peuvent être continuellement créés ou perdus par remaniements génomiques. Il est de ce fait difficile d'être certain d'une évolution commune, un satellite actuel pouvant dériver d'un ancêtre qui, depuis, a disparu. La plus importante caractéristique de ces satellites est qu'ils représentent de très longs rubans d'ADN très peu complexes et dont la séquence reste constante.

Tableau 21.1
Chez les *D. virilis*, les satellites d'ADN sont apparentés.

Satellite	Séquence prédominante	N. copies /génome	% du génome
I	ACAACT TGTTGA	1,1 × 10 ⁷	25 %
II	ATAACT TATTGA	3,6 × 10 ⁶	8 %
III	ACAAAT TGTTAA	3,6 × 10 ⁶	8 %
Cryptique	AATATAG TTATATC		

Plus de 95 % de chaque satellite est dû à la répétition en tandem de la séquence prédominante. On peut noter que les satellites II et III dont la composition nucléotidique est exactement la même (1 paire de G-C sur 7, soit 14 % de G-C) n'ont pas les mêmes densités (1,688 et 1,671 g/cm³ respectivement).

Une autre caractéristique de ces satellites est une importante asymétrie de l'orientation des paires de bases des deux brins. Le **Tableau 21.1** montre que chez « *D. virilis* », l'un des deux brins des satellites principaux est plus riche en nucléotides T ou G. Ceci a pour conséquence d'augmenter sa densité, dans des conditions dénaturantes ; il sera possible alors de séparer ce **brin lourd (H)** de son complémentaire le **brin léger (L)**. Ceci peut être utile pour séquencer le satellite.

LES SATELLITES DES MAMMIFÈRES
SONT COMPOSÉS D'UNITÉS DE RÉPÉTITIONS
HIÉRARCHISÉES

Les études effectuées chez les mammifères, et notamment chez divers rongeurs, ont montré de notables divergences entre les unités répétées en tandem incluses dans chaque satellite. Il est possible par traitement chimique ou enzymatique de séparer parmi les fragments d'oligonucléotides libérés, de courtes séquences plus représentées. Cependant, ces courtes séquences prédominantes ne représentent qu'une petite minorité des copies. Les autres séquences de petite taille, non prédominantes, dérivent de la séquence majoritaire par toutes sortes de substitutions, délétions et insertions.

Cependant, il est possible qu'une succession de ces variants de la petite unité constitue une plus longue unité de répétition, elle-même répétée en tandem avec quelques modifications. C'est ainsi que chez les mammifères,

les ADN satellites s'organisent à partir d'une *hiérarchie* d'unités de répétition. Ces unités de répétition de plus grande taille constituent les séquences qui se renaturent lors des expériences de réassociation. On les caractérise également par digestion à l'aide de certaines enzymes de restriction.

Il est difficile de se faire une idée de la complexité de l'ADN répétitif à cause des conséquences des erreurs d'appariements sur la réaction de réassociation. Celles-ci sont particulièrement sensibles dans l'ADN satellite ; ce qui signifie qu'il ne sera possible d'estimer la longueur de l'unité de réassociation que selon un facteur d'environ deux, au mieux.

Si l'on digère un ADN satellite par une enzyme de restriction possédant un site de reconnaissance dans son unité de répétition, on obtiendra un fragment pour chaque unité renfermant ce site. En fait, lorsque l'on digère l'ADN du génome eucaryote, la plupart de l'ADN est présent dans une traînée, conséquence de la répartition aléatoire des sites de clivage. A l'inverse, l'ADN satellite est résolu en bandes fines, le clivage en des sites régulièrement répartis faisant apparaître un grand nombre de fragments de taille identique ou presque.

Il est peut-être difficile de déterminer la séquence de l'ADN satellite. On peut essayer d'obtenir directement la séquence en utilisant les bandes faibles produites par le clivage. Cependant, bien qu'il existe d'appréciables divergences entre des unités de répétition indépendantes, certains nucléotides sont présents en même position dans différentes unités, rendant ininterprétable le gel de séquence. Si la divergence n'est pas trop élevée, de l'ordre de 2 %, il sera possible de déterminer la séquence moyenne de l'unité de répétition.

Il est possible de cloner les unités de répétition distinctes dans des plasmides. La difficulté est que la séquence satellite a tendance à être excisée du plasmide recombinant par recombinaison dans l'hôte bactérien. Cependant, si le clonage réussit, la séquence du fragment cloné sera alors déterminée sans ambiguïté. Ce procédé ne donnant que la séquence d'une ou plusieurs unités de répétition, il faudra, pour reconstituer le type de diver-

gence caractéristique d'un satellite, connaître la séquence de nombreux fragments indépendants.

Quelle que soit la stratégie utilisée pour le séquençage, la quantité d'informations obtenues sera fonction de la longueur pouvant être analysée à partir d'un gel de séquence. Du fait de la répétition de copies divergentes en tandem, il sera impossible de reconstruire de plus longues séquences par la stratégie de chevauchement de différents fragments de restriction.

L'ADN satellite de la souris « *M. musculus* » est coupé par EcoRII en une série de bandes dont un fragment majoritaire monomérique de 234 pb. On retrouve cette séquence avec quelques variantes tout au long des 60 à 70 % du satellite qui sont coupés en cette bande monomérique. Analysons maintenant cette séquence par ordre de taille décroissante des unités de répétition la composant.

La **Figure 21.3** représente la séquence en deux « demi-répétitions ». Si nous décomposons la séquence de 234 pb en deux éléments de 117 pb et que nous alignons le deuxième élément par rapport au premier, il apparaît que ces deux moitiés sont très homologues. Elles ne diffèrent qu'en 22 positions, ce qui représente un taux de divergence de 19 %. Ceci signifie que la séquence classique de 234 pb a été produite au cours de l'évolution par duplication d'une unité de répétition de 117 pb, après quoi des différences se sont accumulées entre les doublets.

Il est possible de caractériser au sein du fragment de 117 pb deux autres sous-unités. Chacune d'elle est une « répétition d'un quart » du satellite entier. Les quatre « répétitions d'un quart » sont alignées **Figure 21.4**. Les deux lignes supérieures représentent la première « demi-répétition » de la Figure 21.3, les deux lignes inférieures représentent la seconde « demi-répétition ». Nous voyons que les divergences entre les quatre « répétitions d'un quart » ont augmenté : 23 sur 58 positions au total soit 40 %. Il semble que les trois premiers quarts présentent un peu plus d'homologies, une grande proportion des divergences étant due à des variations du quatrième quart.

Figure 21.3

L'unité de répétition de l'ADN satellite de souris contient deux « demi-répétitions » ; leurs différences sont indiquées en couleur.

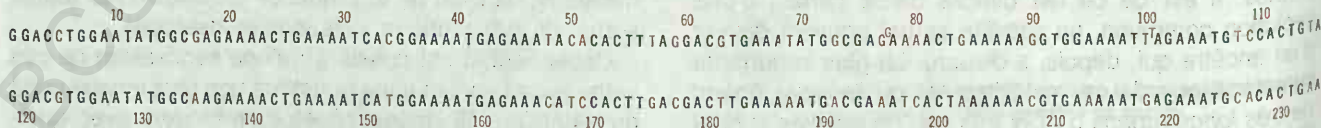


Figure 21.5
L'alignement des « répétitions d'un huitième » montre que chaque « répétition d'un quart » est constituée de deux moitiés α et β . La séquence consensus est fournie par la base la plus représentée à chaque position. La séquence ancestrale se décompose en trois unités de 9 pb très voisines de la séquence consensus et qui pourraient être à l'origine des sous-unités α et β . (Noter que la séquence d'un satellite est continue et doit être considérée comme circulaire si l'on veut définir une séquence consensus. C'est ce qu'indique le pointillé joignant le dernier triplet GAA aux premières 6 pb.)

1	2	3	4	5	6	7	8	9	
			G	G	A	C	C	T	
G	G	A	A	T	A	T	G	G	C
G	A	G	A	A	A	A	C	T	
G	A	A	A	A	T	C	A	C	
G	G	A	A	A	A	T	G	A	
G	A	A	A	T*	C	A	C	T	
T	T	A	G	G	A	C	G	T	
G	A	A	A	T	A	T	G	G	C
G	A	G	G ^A	A	A	A	C	T	
G	A	A	A	A	A	G	G	T	
G	G	A	A	A	A	T	T	A	
G	A	A	A	T*	C	A	C	T	
G	T	A	G	G	A	C	G	T	
G	G	A	A	T	A	T	G	G	C
A	A	G	A	A	A	A	C	T	
G	A	A	A	A	T	C	A	T	
G	G	A	A	A	A	T	G	A	
G	A	A	A	C*	C	A	C	T	
T	G	A	C	G	A	C	T	T	
G	A	A	A	A	A	T	G	A	C
G	A	A	A	T	C	A	C	T	
A	A	A	A	A	A	C	G	T	
G	A	A	A	A	A	T	G	A	
G	A	A	A	T*	C	A	C	T	
G	A	A							

* indique un triplet inséré dans la séquence β

Le C en position 10 est une base supplémentaire de la séquence α

G₂₀ A₁₆ A₂₁ A₂₀ A₁₂ A₁₇ T₈ G₁₁ A₅
 T₇ C₅ A₈ C₉ T₁₅
 C₇

Figure 21.6
L'existence d'une séquence consensus globale apparaît en écrivant la séquence satellite selon une répétition de 9 pb.

A l'intérieur de la séquence consensus, on peut identifier trois séquences apparentées de 9 pb. Si dans une des répétitions nous considérons la base suivante la plus fréquente à deux positions de cette répétition, nous pouvons en effet la diviser en trois séquences voisines de 9 pb très bien apparentées.

GAAAAACGT
GAAAAATGA
GAAAAAACT

C'est une des meilleures conjectures que nous puissions faire concernant la constitution originelle du satellite.

La séquence consensus est directement analysée à la **Figure 21.6** ; elle démontre que la séquence satellite actuelle peut être considérée comme dérivée d'une séquence de 9 pb. La séquence consensus globale est : GAAAAA^{TC}_{AG} T, qui est effectivement un amalgame de la **Figure 21.5**.

EN RETRAÇANT LES STADES D'ÉVOLUTION DE L'ADN SATELLITE DE SOURIS

D'après la séquence de l'ADN satellite nous pouvons retracer son évolution et estimer ses propriétés actuelles. La **Figure 21.7** présente un modèle montrant les étapes probables de son évolution.

Ce modèle est régi par le principe général selon lequel, à des moments divers, un groupe d'unités de répétitions peut avoir brusquement subi une amplification latérale, engendrant un grand nombre de copies en tandem identiques. On appelle un tel événement **réplication par sauts**. Ensuite les séquences des copies divergent au fur et à mesure qu'elles accumulent les mutations. Ultime-riement, un groupe de ces copies pourra lui-même être l'objet d'une nouvelle phase de réplication par saut. L'étendue de la divergence dans les copies amplifiées à chaque « saut » dépendra du temps écoulé depuis la dernière réplication par saut ayant imposé une identité au satellite. Le satellite peut évoluer par une série de ces réplications par sauts successifs alternant avec une accumulation de mutations.

Supposons que le satellite actuel trouve son origine dans la répétition en tandem d'une séquence telle que GAAAAATGT ou quelque chose d'approchant. (Cette séquence pourrait avoir fait partie d'un satellite ou alors avoir une origine différente : son existence remonte aussi loin qu'il est possible de suivre le satellite.) Toutes les unités originelles de 9 pb étaient identiques, mais avec le temps, les mutations ont créé des différences entre elles. Puis trois unités adjacentes ayant la séquence ancestrale dont nous avons fait l'hypothèse à la **Figure**

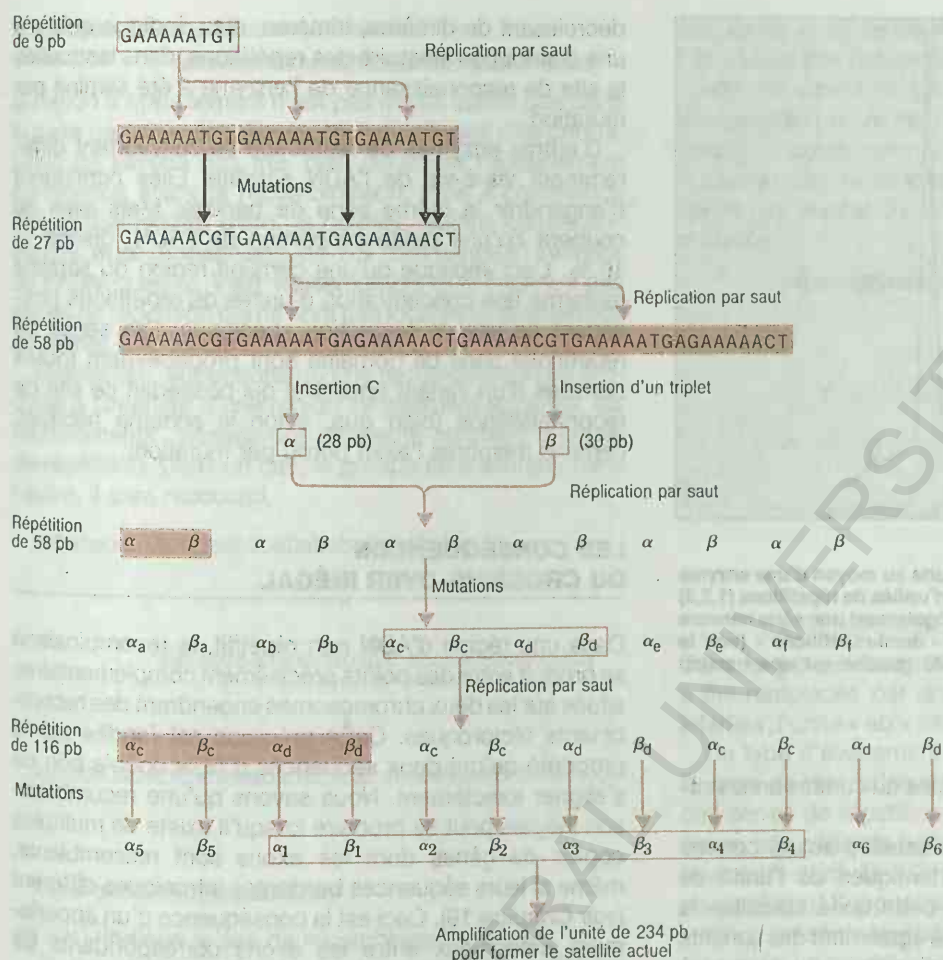


Figure 21.7
L'évolution de l'ADN satellite de souris s'explique par une alternance de réplifications par saut et de mutations accumulées.

21.5 ont été amplifiées et ont donné une répétition en tandem de 27 pb.

Des mutations se sont produites dans cette unité ; dans certains cas, une unité a gagné un C supplémentaire pendant que sa voisine se voyait insérer un triplet. Cette paire de répétitions, qui, à elles deux, sont longues de 58 pb, a subi une réplification par saut aboutissant à un satellite que nous pouvons définir comme $(\alpha\beta)_n$.

A nouveau, le satellite a accumulé des mutations ponctuelles, des délétions et des insertions, qui créeront une divergence entre ses unités de répétition. Deux paires de bases $\alpha\beta$ adjacentes ont été utilisées dans une réplification ultérieure par saut donnant une unité de répétition ($\alpha_c \beta_c \alpha_d \beta_d$) de 116 pb. Après de nouvelles mutations, deux de ces unités adjacentes ont été amplifiées pour donner l'actuel satellite.

LES VARIATIONS DANS L'ACTUELLE UNITÉ DE RÉPÉTITION

La séquence moyenne du fragment monomère de l'ADN satellite de souris explique ses propriétés. L'unité répétée la plus longue qui mesure 234 pb est identifiée par sa carte de restriction. L'unité de réassociation entre simples brins de l'ADN dénaturé du satellite est probablement une « demi-répétition » de 117 pb, car les fragments de 234 pb peuvent s'associer à la fois sur toute leur longueur, ou sur la moitié (dans ce dernier cas, la première « demi-répétition » est renaturée avec la seconde « demi-répétition » de l'autre). Dans les produits de digestion de l'oligonucléotide, la majorité des fragments communs, représentant 4 % de la quantité totale d'ADN, sont GA_5TGA , GA_4TGA , et GA_4CTGA . Ils sont présents dans

xabababababababababababababy

xabababababababababababababababy

X

xabababababababababababababy

↓

xabababababababababababababababababy

+

xabababababababababy

Mais supposons que les composants « a » et « b » de l'unité de répétition soient eux-mêmes suffisamment bien apparentés pour s'apparier. Les deux groupements peuvent alors s'aligner en **contre-phase** avec la séquence « a » de l'un, alignée sur la séquence « b » de l'autre. La

xabababababababababababy

X

abababababababababababy

↓

xababababababababa**abababy**

+

xababababababababbababy

En retournant l'argument, l'identification de la série de demi-répétitions sur le gel nous montre que l'unité répétée de 234 pb est constituée de deux demi-répétitions assez bien apparentées pour parfois s'apparier par recombinaison. Nous pouvons également voir dans la Figure 21.8 quelques bandes plus faibles correspondant à des inter-

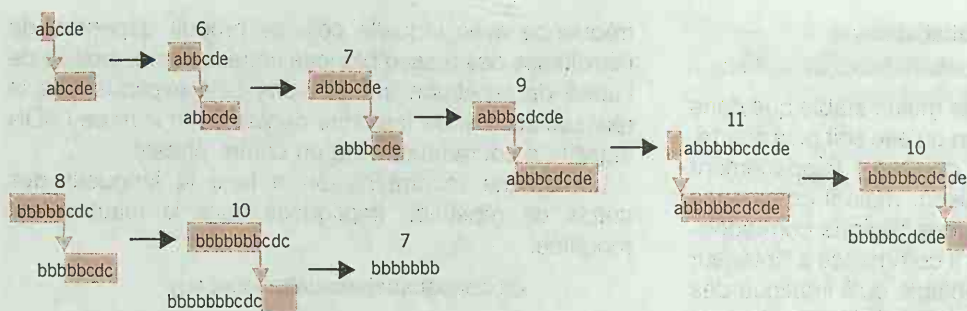


Figure 21.9

Une recombinaison inégale permet à une unité de répétition donnée d'occuper entièrement le groupe. Les chiffres indiquent la longueur de l'unité de répétition à chaque stade.

valles d'1/4 et de 3/4. Ces derniers seront créés de la même manière que les demi-intervalles lorsque la recombinaison a lieu entre des groupes alignés sur un quart de la longueur de l'unité entière. Les parentés décroissantes entre les répétitions d'un quart et les demi-répétitions expliquent la réduction de la fréquence des bandes d'1/4 et 3/4 par rapport aux bandes d'1/2.

LA FIXATION PAR RECOMBINAISON POURRAIT PERPÉTUER DES RÉPÉTITIONS IDENTIQUES

Nous présumons (généralement) que la séquence du satellite d'ADN n'est pas exposée à une forte pression sélective (si pression il y a). Contrairement à une séquence codant pour une protéine, où une mutation peut neutraliser le produit, il est vraisemblable que le satellite d'ADN doive sa fonction à sa seule présence et non à sa séquence exacte.

Cette idée correspond parfaitement à la structure même des satellites dont les unités de répétition sont plus apparentées qu'identiques. Des mutations se sont accumulées depuis que le satellite a été réarrangé pour la dernière fois en une séquence uniforme. Mais comment allons-nous expliquer la présence chez les arthropodes, de satellites dont la plupart des unités de répétition restent identiques ? Même si les séquences étaient importantes il resterait encore la difficulté de savoir comment la sélection naturelle peut être imposée sur de si nombreuses copies.

Nous avons proposé la recombinaison inégale comme alternative à la réplication par saut pour expliquer l'évolution de l'ADN satellite. L'idée de base est qu'un crossing-over inégal se produit fréquemment dans des sites aléatoires. Une série d'événements aléatoires fait qu'une unité de répétition prend le dessus sur le satellite entier. Ce processus est appelé **fixation par recombinaison**.

La diffusion d'une unité de répétition spécifique à travers le satellite est illustrée par la **Figure 21.9**. Supposons qu'un satellite soit initialement constitué d'une séquence *abcde*, où chaque lettre représente une unité de répétition. Les différentes unités de répétition sont suffisamment similaires les unes aux autres pour s'apparier de façon aberrante lors d'une recombinaison.

Ce modèle nous permet l'extrapolation suivante : toute séquence d'ADN, qui n'est pas sous pression sélective, sera supplantée par une série de répétitions en tandem identiques produites de cette façon. L'hypothèse décisive est que le processus de fixation par recombinaison * est plus rapide que les mutations, de sorte que les nouvelles mutations sont, soit éliminées (leurs répétitions sont perdues), soit viennent à remplacer le groupe entier (comme le montre la diffusion du variant « b » dans la Figure 21.9).

Nous pouvons nous demander si l'existence de domaines renfermant des unités de répétition spécifiques (comme dans l'ADN satellite de la souris) pourrait représenter une diffusion partielle de l'unité depuis son origine. Par exemple, nous pouvons voir que dans les stades intermédiaires de la diffusion illustrés par la **Figure 21.9**, il existe des groupes, dont le domaine est constitué d'une série de variants « b ».

POUR EN SAVOIR PLUS

L'idée d'une réplication par saut a été introduite par Southern (*J. Mol. Biol.* 94, 51-70, 1975). Les implications du crossing-over inégal ont été développées par Smith (*Science* 191, 528-535, 1976). L'organisation des répétitions inversées et des répétitions en tandem a été passée en revue dans le *Gene Expression* de Lewin, 2, *Eucaryotic Chromosomes* (Wiley, New York, 1980, pp. 531-569).

* N. d. T. : Un phénomène d'homogénéisation des séquences d'une famille d'éléments répétés, et donc tout à la fois de relative fixation de ces séquences et de leur évolution concertée, peut formellement être expliqué, soit par des processus de recombinaison inégale comme indiqué par l'auteur, soit par processus de conversion génique qui n'entraîne pas alors le phénomène d'expansion-réduction du groupe répété.

CHAPITRE 22

EXCISION ET RABOTAGE

PARTIE 7

VERS LE PRODUIT FINAL : LA MATURATION DES ARN

Le code programme doit, par lui-même, être le facteur principal contrôlant le développement [de l'organisme]. Mais... l'image moléculaire du gène rend concevable l'existence d'un code secondaire correspondant avec précision à un plan de développement spécifique et extrêmement compliqué, contenant en quelque sorte les moyens de sa mise en œuvre.

Erwin Schrödinger, 1945

BCU IASI / CENTRAL UNIVERSITY LIBRARY

PARTIE 7

LA MATURATION DES ARBRES

Le processus de maturation des arbres est un phénomène complexe qui implique de nombreux facteurs, tels que la température, l'humidité, la lumière et la composition chimique du milieu. Ces facteurs agissent en synergie pour déclencher et réguler les réactions biochimiques responsables de la maturation. Les arbres subissent des changements significatifs au niveau de leur structure et de leur composition chimique au cours de ce processus, ce qui leur permet de passer d'un état immature à un état mature capable de résister aux stress environnementaux et de se reproduire.

CHAPITRE 22

EXCISION ET RABOTAGE D'UN ARN STABLE

La **maturation** d'une molécule à partir de son précurseur représente un processus décisif dans la production de nombreuses formes d'ARN. Les ARNr et les ARNt procaryotes et eucaryotes sont synthétisés sous la forme de transcrits primaires plus complexes à partir desquels doit être dégagée la molécule d'ARN mature. Inversement, la plupart des ARNm procaryotes *sont* des transcrits primaires ; donc la maturation n'est nécessaire à la traduction que dans des cas exceptionnels. Cependant, chez les eucaryotes, la production d'un ARNm à partir d'un gène morcelé constitue la partie la plus intensive du travail de toute la maturation de l'ARN ; elle peut impliquer une série de réactions d'excision-épissage qui raccordent les exons entre eux en supprimant les introns (voir le Chapitre 23).

Les réactions de la maturation sont extrêmement spécifiques, produisant des molécules d'ARN matures aux extrémités 5' et 3' particulières. Des ribonucléases spécifiques sont responsables de ces réactions : elles dépendent de la capacité enzymatique à rompre, dans l'ARN, les liaisons phosphodiester. Comme pour toutes les nucléases, les enzymes sont regroupées en deux catégories principales : les **exonucléases** et les **endonucléases**.

Les endonucléases coupent les liaisons individuelles à l'intérieur des molécules d'ARN, engendrant des fragments discontinus. Elles sont impliquées dans les réactions de coupure lorsque la séquence mature doit être séparée de la séquence bordante.

Les exonucléases éliminent les résidus, un par un, à partir de l'extrémité de la molécule, engendrant des mononucléotides. Leur attitude à l'égard du substrat peut être une attitude **aléatoire** ou bien **ordonnée**. Une enzyme agissant aléatoirement supprime une base dans une molécule d'ARN, puis s'en désintéresse ; pour sa prochaine action catalytique, l'enzyme devra attaquer une molécule différente d'ARN substrat. L'action ordonnée signifie que l'enzyme reste liée à une molécule de substrat, supprimant base après base jusqu'à ce que sa mission soit accomplie.

Les exonucléases sont impliquées dans les réactions de rabotage. Dans la perspective du substrat ARN, les résidus excédentaires sont peu à peu « grignotés », base par base. Toutes les exonucléases connues se déplacent le long de la chaîne d'acide nucléique en partant de l'extrémité 3'. *Ainsi le matériel excédentaire présent à l'extrémité 3' peut être raboté, mais le matériel excédentaire obtenu à l'extrémité 5' ne peut être libéré que par coupure.*

Lorsque le pont phosphodiester qui relie deux nucléotides est coupé, il peut l'être en principe de n'importe quel côté du groupement phosphate. Les conséquences en sont illustrées par la **Figure 22.1**. La coupure sur l'un des deux côtés engendre des terminaisons 3' hydroxyle et 5' phosphate. La coupure sur l'autre côté engendre des terminaisons 3' phosphate et 5' hydroxyle.

Deux ribonucléases offrant un intérêt pratique, bien que de fonction physiologique encore inconnue, sont la

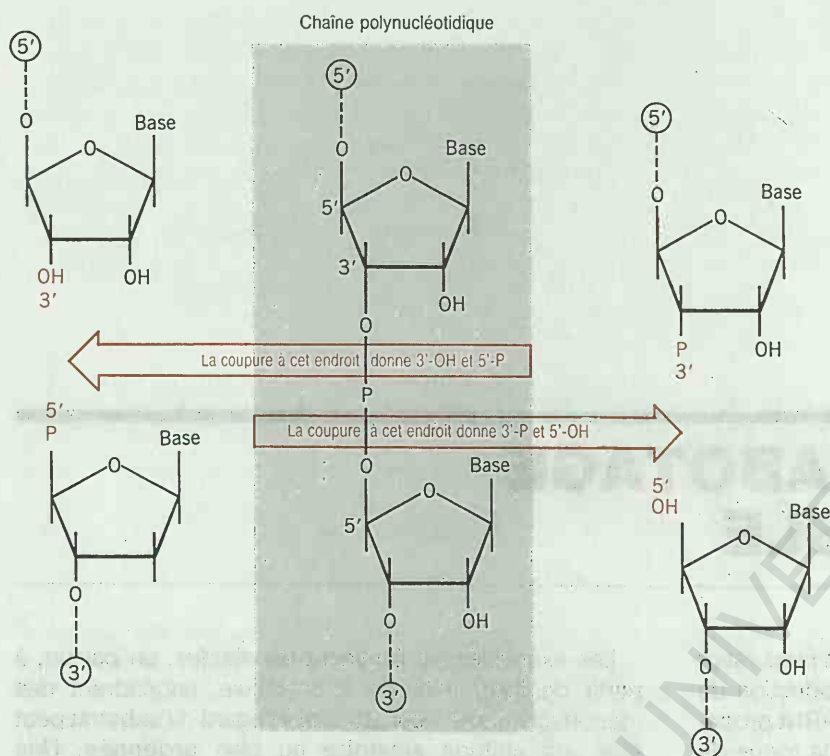
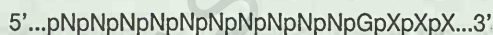


Figure 22.1

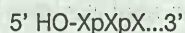
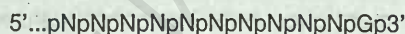
La nature des nouvelles extrémités est déterminée par le côté où la liaison phosphodiester est coupée.

ribonucléase T1 et la ribonucléase pancréatique. Avant l'apparition des méthodes de routine pour le séquençage de l'acide nucléique, ces enzymes étaient utilisées pour digérer l'ARN en oligonucléotides, à partir desquels la séquence de la molécule pouvait ensuite être reconstituée. Ces deux ribonucléases sont des endoribonucléases reconnaissant des séquences relativement simples.

La ribonucléase T1 est commune dans les champignons. Elle hydrolyse la liaison située en 3' d'un résidu G. Ainsi elle coupe une séquence telle que :



pour engendrer les produits :



Notons que les nouvelles extrémités ont des terminaisons 3' phosphate et 5' hydroxyle. La ribonucléase pancréatique engendre des extrémités de même nature et coupe en 3' de n'importe quelle pyrimidine.

Ces extrémités sont totalement différentes des types

de terminaisons rencontrées dans les ARN naturels, qui ont un phosphate (ou un phosphate modifié) à l'extrémité 5', et un hydroxyle à l'extrémité 3'. Les enzymes impliquées dans la maturation de l'ARN engendrent des extrémités similaires (5' phosphate et 3' hydroxyle), exception faite des nucléases impliquées dans l'excision-épissage où des extrémités inhabituelles peuvent être produites (voir Chapitre 23).

Les ribonucléases ont des spécificités considérablement variables. Aucune, à notre connaissance, ne requiert pour substrat des séquences de bases multiples précises. Ainsi il n'existe pas d'équivalent des nucléases de restriction qui, dans l'ADN, attaquent des séquences spécifiques. De nombreuses ribonucléases ont une spécificité réduite, en ce sens qu'elles coupent les liaisons phosphodiester au niveau de bases particulières.

Les ribonucléases non seulement prennent part aux réactions spécifiques de la maturation, mais sont aussi responsables de la dégradation de l'ARN superflu. Les molécules « superflues » comprennent des espèces

matures en fin de course, ainsi que du matériel étranger éliminé pendant la maturation des précurseurs. En particulier, les ARN messagers qui ont fait leur temps doivent être dégradés (le renouvellement est rapide dans la bactérie, plus lent dans les eucaryotes, mais l'ARNm n'est assez stable pour survivre pendant une période prolongée que dans de très rares cas).

Nous pourrions nous attendre à ce que les ribonucléases impliquées dans le renouvellement de l'ARN soient moins spécifiques que celles impliquées dans des étapes précises de la maturation, puisque leur rôle est davantage de dégrader totalement l'ARN que d'engendrer des terminaisons spécifiques. Dans cette catégorie on trouve à la fois les endoribonucléases et les exoribonucléases. La dégradation d'un ARN en mononucléotides constitue sa maturation finale, restituant ses composants en vue d'une réutilisation par la voie métabolique appropriée.

Les enzymes possédant une plus grande spécificité peuvent reconnaître la conformation du substrat ARN, plutôt qu'une séquence de bases particulières. L'identification d'un substrat peut prendre la forme de la recherche d'une caractéristique spécifique de la structure secondaire (comme une épingle à cheveux d'une certaine taille), ou bien d'un examen d'ensemble de la structure secondaire/tertiaire (comme cela semble être le cas dans la maturation de l'ARNt).

Pour comprendre l'afflux et le reflux des nucléotides dans l'ARN, nous aimerions pouvoir caractériser l'ensemble des ribonucléases d'un type donné de cellules. Le seul cas où cela paraisse possible est représenté par *E.*

coli. Dans ce cas il est non seulement possible de caractériser biochimiquement les enzymes, mais également de déterminer leurs effets *in vivo* en sélectionnant des bactéries dans lesquelles une enzyme donnée a été mutée. En effet, ces deux catégories de preuves sont souvent nécessaires pour distinguer des enzymes, ou pour montrer que deux activités a priori différentes peuvent être attribuées à une même enzyme. (Une semblable approche pourrait d'ici peu s'appliquer à la levure.)

Nous ne savons pas encore précisément combien de ribonucléases renferme *E. coli*, mais le chiffre actuel est d'environ 10. Les enzymes ayant été purifiées sont énumérées dans le **Tableau 22.1**. Il semble probable que diverses combinaisons d'un petit nombre d'activités des ribonucléases soit suffisant pour assurer la maturation de l'ARNr et de l'ARNt. Bien que l'ARNm soit activement dégradé, sa maturation a été difficile à analyser jusqu'à présent, et nous ne possédons aucune information détaillée sur les activités enzymatiques qui en sont responsables.

Ces ribonucléases jouent-elles des rôles essentiels dans la cellule ? Des formes mutantes de certaines endoribonucléases (mise à part la ribonucléase I qui est sans effet) accumulent des précurseurs immatures, mais sont viables. Les mutants de certaines exonucléases possèdent souvent des phénotypes apparemment intacts, ce qui pourrait signifier que si une enzyme vient à manquer, une autre peut se substituer à elle. Les mutants ayant perdu un grand nombre d'enzymes ne sont pas toujours viables.

Tableau 22.1
E. coli contient un nombre relativement faible de ribonucléases.

Enzyme	Locus	Classe de l'ARN substrat	Type
Ribonucléase P	<i>mpA</i>	ARNt, extrémité 5'	endo
»	<i>mpB</i>		
Ribonucléase BN	?	ARNt, extrémité 3'	exo
Ribonucléase D	<i>rnd</i>	ARNt, extrémité 3'	exo
Ribonucléase T	?	ARNt, CCA 3'	exo
Ribonucléase II	<i>rnc</i>	ARNr & ARNm	endo
Ribonucléase R	?	ARNr & ARNm	exo
Ribonucléase E	<i>rne</i>	ARNr 5S	endo
Ribonucléase I	<i>rna</i>	la plupart des ARN	endo
Ribonucléase II	<i>rnb</i>	ARN, non-structuré en 3'	exo
Polynucléotide phosphorylase	<i>pnp</i>	ARN, non-structuré en 3'	exo
Ribonucléase H	<i>mh</i>	hybrides ARN-ADN	endo

L'ARNase III LIBÈRE LES ARNm PRÉCOCES DU PHAGE T7

Les ARNm bactériens sont pratiquement tous traduits sous la forme dans laquelle ils sont transcrits. La plupart sont polycistroniques, bien que certains soient monocistroniques. Les gènes précoces du phage T7 présentent néanmoins une situation exceptionnelle. Ils sont transcrits sous forme d'un ARNm polycistronique. La molécule est ensuite coupée en différents sites, ce qui libère les ARNm de chaque gène. (Un découpage similaire peut également se produire pour l'ARN tardif, mais il est moins bien défini.)

On trouve cette forme d'expression chez le phage T3 comme chez le phage T7. Comme les deux séquences d'ADN des phages sont relativement divergentes, la survivance d'un processus commun suggère que celui-ci offre certains avantages sélectifs. Mais nous ignorons pourquoi ces deux phages ont sélectionné ce processus particulier de maturation de l'ARNm.

La transcription de la région précoce du phage T7 est illustrée à la **Figure 22.2**. L'initiation a lieu dans un des groupes des trois promoteurs (A1, A2, A3) et se poursuit jusqu'à un terminateur (*t*) environ 7 000 paires de bases plus loin. Six gènes sont localisés dans cette région, et l'**ARNase III** coupe le transcrit aux frontières intercistroniques pour libérer cinq ARNm indépendants (quatre monocistroniques et un bicistronique).

Quelle est la nature des sites reconnus par l'ARNase III ? Les sites cibles ont été identifiés en établissant les séquences des extrémités 5' et 3' des molécules d'ARNm engendrées par le clivage. Il est possible d'aligner ces séquences par rapport à la séquence connue d'ADN du phage T7, afin de mettre en évidence l'environnement complet du site de clivage.

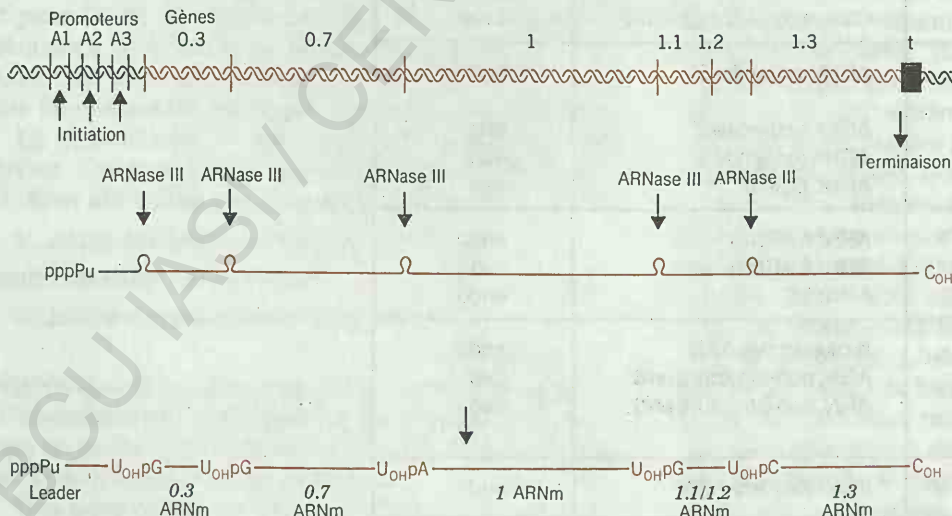
Trois des sites précoces de coupure du T7 ont des séquences présentant une importante homologie, mais les deux autres sont différents. La plus longue homologie commune aux cinq sites est un triplet proche du site de coupure. La seule caractéristique connue commune à tous les sites de coupure de l'ARNase III est la présence d'une structure en double brin. Tous les sites du T7 prennent la forme d'un double brin en épingle à cheveux contenant une « bulle » non appariée. Un exemple en est tracé à la **Figure 22.3**.

L'ARNase III coupe en un point unique dans quatre des sites et en deux points dans le site restant, à l'intérieur ou près de la « bulle ». La seule information dont nous disposons quant aux interrelations entre la coupure et la structure de l'ARN est que des mutations réduisant la stabilité de l'épingle à cheveux peuvent diminuer ou supprimer la coupure par l'enzyme.

Dans les cellules mutantes *rnc⁻*, déficientes en ARNase III, l'ADN du phage T7 est transcrit sous la forme d'un grand précurseur unique. Bien que le précurseur ne soit

Figure 22.2

La région précoce du phage T7 est transcrite en un ARN unique, coupé par l'ARNase III en cinq ARNm indépendants.



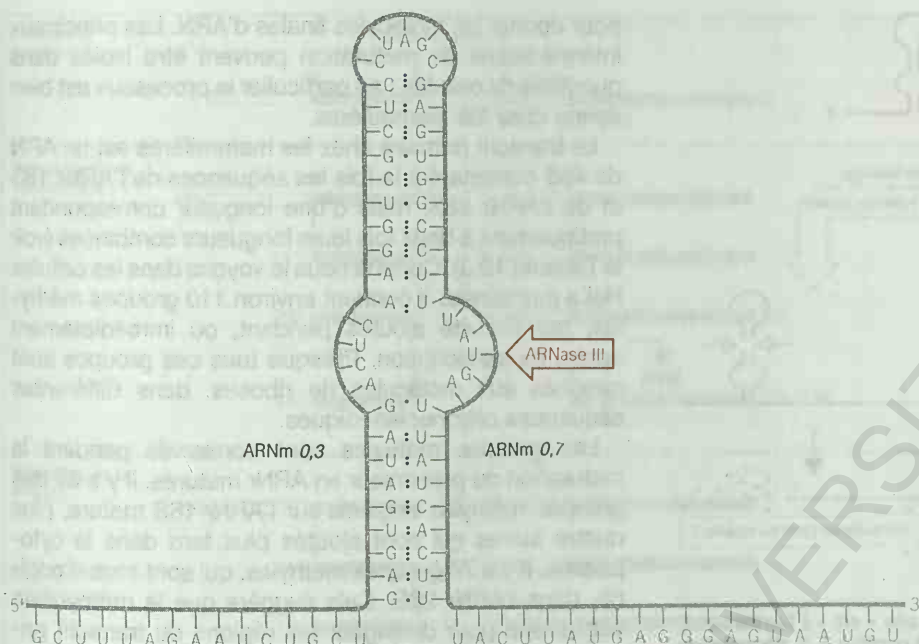


Figure 22.3

Un site de coupeure typique de l'ARNase III dans l'ARN du T7 est situé à l'intérieur d'une courte épingle à cheveux renfermant une « bulle » non appariée.

pas coupé dans un hôte bactérien mutant, il peut être traduit en un ARNm polycistronique, produisant les protéines précoces habituelles, maintenant ainsi l'infection. La réussite de l'infection par le T7 dans les mutants *mnc* nous montre que la réaction de coupeure n'est pas nécessaire à l'expression précoce. Cependant, la coupeure à l'extrémité du gène 1,2 se révèle nécessaire à la traduction de l'ARNm bicistronique 1,1/1,2.

Lorsque cette coupeure est impossible, aucune des deux régions 1,1 ou 1,2 ne peut être traduite. La raison en est que l'ARNm, plus long, présente une structure secondaire qui bloque l'initiation de la séquence codante 1,1 (l'impossibilité de traduire la séquence codante 1,1 empêche à son tour la traduction de la séquence 1,2).

Cet effet a été révélé grâce à des tentatives de culture d'un T7 dans une souche mutante d'*E. coli*, où l'expression du gène 1,2 est indispensable à la croissance du phage (ce n'est habituellement pas le cas). Dans cette situation, et peut-être dans un habitat naturel, le clivage visant à produire une extrémité normale pourrait être une réaction nécessaire.

Un autre exemple dans lequel l'ARNase III pourrait affecter la traduction est fourni par un ARNm du phage

lambda qui porte la séquence du gène *int*. Généralement cet ARNm n'est pas traduit (la protéine *Int* est synthétisée à partir d'un autre ARNm). Des mutations dans un site appelé *sib*, situé en aval de la séquence codante *int*, permettent à la protéine *Int* d'être traduite à partir de l'ARNm habituellement inactif.

La mutation *sib* empêche l'ARNase III de reconnaître un site-cible. La Figure 22.4 montre que la coupeure de l'ARNm de type sauvage en ce site libère probablement l'extrémité 3', en vue de sa dégradation par d'autres ribonucléases (inconnues). L'instabilité de l'ARNm rend impossible la synthèse de la protéine *Int*. Les mutations situées soit dans *sib*, soit dans *mnc* ont le même effet ; elles rendent impossible la coupeure, et donc stabilisent l'ARNm et permettent à la traduction de se produire. Parce que l'effet contrôle l'expression du gène en amont du site où se produit l'interaction en question, on l'appelle **rétrorégulation**.

Dans les deux gènes 1,1/1,2 du T7 et dans le système « *int* » du lambda, il existe une relation entre la capacité de l'ARN à être traduit et sa coupeure par l'ARNase III en un site spécifique. Dans ces deux cas, le mécanisme probable de la relation repose sur l'effet de la coupeure

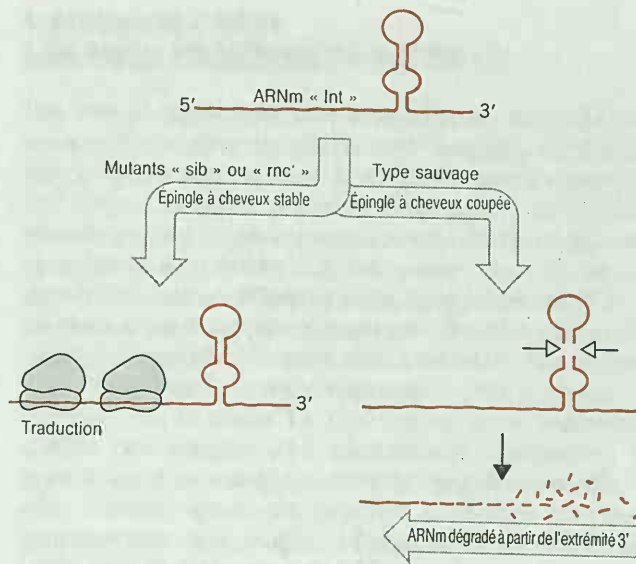


Figure 22.4

La rétrorégulation résulte du clivage du site « *sib* » à l'extrémité 3' de l'ARNm « Int » par l'ARNase III ; cela conduit à la dégradation de l'ARNm.

sur la structure secondaire de l'ARN. En libérant une partie de la molécule, le phénomène de la coupure modifie la structure d'autres régions, dont des sites sont importants pour les fonctions de l'ARN.

Parmi les fonctions affectées, on trouve l'initiation de la traduction, la sensibilité aux dégradations, la capacité à être épissé (chez les eucaryotes : voir le chapitre suivant) et sans nul doute d'autres phénomènes. Nous avons vu au Chapitre 11 que des structures secondaires alternatives interviennent dans l'achèvement de la transcription, constituant de ce fait un réseau régulateur atténuateur. La relation étroite entre la structure secondaire (et peut-être également tertiaire) et la fonction de l'ARN est un thème d'une importance croissante à mesure que s'étend notre connaissance du champ des activités dont l'ARN est capable (voir chapitre suivant).

LES CLIVAGES SONT NÉCESSAIRES POUR LIBÉRER LES ARNr PROCARYOTES ET EUCARYOTES

Nous pouvons suivre la maturation de l'ARNr eucaryote en détail, car le transcrit primaire connaît une maturation relativement lente passant par des étapes intermédiaires

pour donner les molécules finales d'ARN. Les principaux intermédiaires de maturation peuvent être isolés dans quantités de cellules : en particulier le processus est bien connu chez les mammifères.

Le transcrit primaire chez les mammifères est un ARN de 45S contenant à la fois les séquences de l'ARNr 18S et de l'ARNr 28S, mais d'une longueur correspondant pratiquement à deux fois leurs longueurs combinées (voir le Tableau 19.3). Comme nous le voyons dans les cellules HeLa (humaines), il contient environ 110 groupes méthyles, qui ont été ajoutés pendant, ou immédiatement après, sa transcription. Presque tous ces groupes sont attachés aux molécules de riboses, dans différentes séquences oligonucléotidiques.

Les groupes méthyles sont *conservés* pendant la maturation du précurseur en ARNr matures. Il y a 39 des groupes méthyles originels sur l'ARNr 18S mature, plus quatre autres qui sont ajoutés plus tard dans le cytoplasme. Il y a 74 groupes méthyles, qui sont tous d'origine, dans l'ARNr 28S. Cela suggère que la méthylation est utilisée pour distinguer les régions du transcrit primaire qui subissent la maturation en ARNr.

Les séquences des événements de maturation ne sont probablement pas uniques comme l'indiquent les tailles des molécules d'ARN intermédiaires. Mais toutes ces séquences d'événements peuvent être expliquées si l'on suppose qu'un petit nombre de sites de coupure peut être utilisé dans des ordres variés.

Deux processus chez les mammifères, à l'origine caractérisés dans les cellules HeLa (humaines) et L (de la souris) sont illustrés à la Figure 22.5. Dans chaque cas, on trouve des sites de coupure sur le côté 5' du gène 18S, dans la zone intermédiaire entre le gène 18S et le gène 5,8S, et dans l'espace entre les séquences 5,8S et 28S. (La séquence 5,8S se transforme en un petit ARN s'associant à l'ARNr 28S par appariement de bases).

La différence entre les deux processus que nous montre la figure repose uniquement sur l'ordre dans lequel sont utilisés les sites de coupure. D'autres variations sont également possibles. Dans certains cas, on trouve plus d'un processus pour un type donné de cellules.

La figure nous montre le nombre minimum de sites de coupure potentiels utilisés pour engendrer toutes les extrémités 5' et 3'. Nous ne savons pas si ces coupures engendrent réellement les extrémités matures, ou bien si elles libèrent des précurseurs individuels qui sont ensuite coupés et rabotés. Nous ne possédons non plus aucune information relative aux activités des ribonucléases responsables de la maturation. Cependant, nous savons que

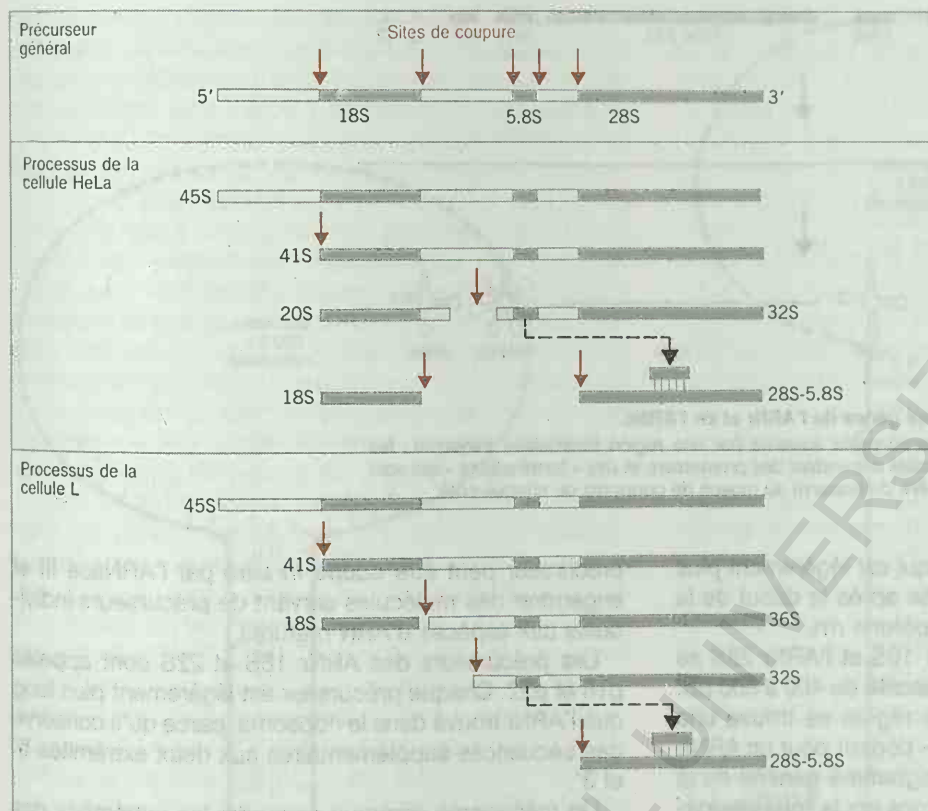


Figure 22.5
Le précurseur 45S de mammifère libère les ARNr matures par clivage dans les régions intercalaires.

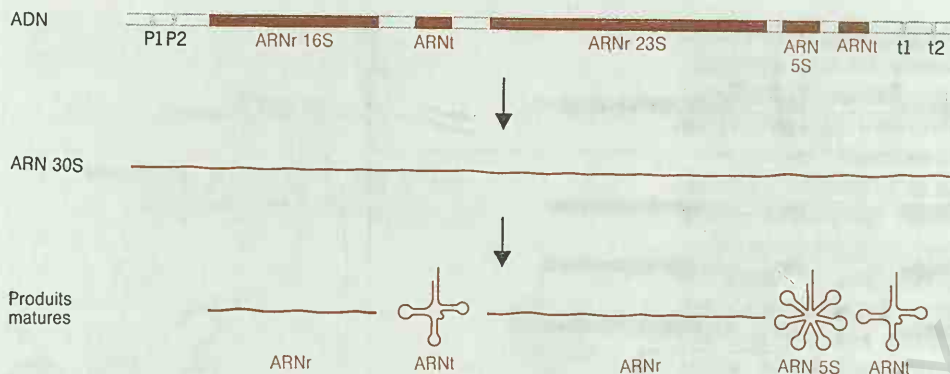
l'ARN 45S s'associe à des protéines dès sa synthèse, de sorte que le processus de maturation vise une ribonucléoprotéine plutôt qu'un ARN libre.

Des coupures sont également nécessaires pour libérer les ARNr bactériens du précurseur. Les sept opérons codant pour l'ARNr dans *E. coli* sont appelés *rrnA-G*. Ils ne sont pas étroitement liés sur le chromosome et nous n'avons aucune information nous permettant de savoir comment l'homogénéité de la séquence est maintenue. Le nombre de gènes étant faible, il devrait être possible de déterminer leurs séquences exactes, et ainsi de voir directement si tous les gènes d'ARNr sont identiques ou bien s'ils possèdent des variations mineures non décelables dans la population d'ARNr.

Chacun de ces opérons est transcrit sous la forme d'un précurseur unique ARN duquel sont libérées, par coupure, les molécules d'ARNr matures. Lorsque la réaction de

coupure est bloquée, le précurseur s'accumule sous la forme d'un ARN sédimentant à environ 30S. L'organisation globale de tous les opérons *rrn* est la même ; ils contiennent tous, les trois molécules d'ARNr dans l'ordre 16S-22S-5S, comme nous le montre la **Figure 22.6**. Les deux ARNr principaux sont situés dans le même ordre que dans l'unité de transcription eucaryotique ; la présence de l'ARN 5S dans la même unité de transcription est propre aux procaryotes.

Tous les opérons *rrn* ont une double structure promotrice. Le premier promoteur, *P1*, est situé à environ 300 pb en amont du début de la séquence de l'ARNr 16S. C'est probablement le promoteur principal. Jusqu'aux premières 150 pb de l'unité de transcription, il peut exister des variations dans les différents opérons *rrn*. A l'intérieur de cette région, à peu près à 110 pb de *P1*, se trouve un second promoteur, *P2*. L'extrémité 5' du

**Figure 22.6**

Les opérons *rrn* contiennent à la fois les gènes de l'ARNr et de l'ARNt.

Chaque séquence fonctionnelle est séparée de la suivante par une région intercalaire transcrite ; les longueurs des séquences leader et terminales dépendent des promoteurs et des « terminateurs » qui sont utilisés. Chaque produit ARN doit être libéré du transcrit au moyen de coupures de chaque côté.

précurseur propre à l'ARNr 16S (qui est légèrement plus long que l'ARNr mature) est située après le début de la séquence commune à tous les opérons *rrn*.

Entre les séquences de l'ARNr 16S et l'ARNr 22S se trouve une région intercalaire transcrite de 400 à 500 pb. Cependant, à l'intérieur de cette région se trouve une séquence — ou des séquences — codant pour un ARNt. Comme il est résumé dans le diagramme général de la Figure 22.6, dans quatre des opérons *rrn*, la zone intercalaire transcrite renferme une séquence unique d'ARNt, celle de l'ARNt^{Glu}. Dans les trois autres opérons *rrn*, la zone intermédiaire transcrite renferme les séquences de deux ARNt (l'ARNt^{Leu} et l'ARNt^{Arg}). Ainsi, sur la base des séquences des « ARNt intercalaires », nous pouvons définir deux types d'opérons d'ARNr. Les ARNt sont libérés de l'ARN précurseur par coupure ; les régions de la zone intermédiaire qui restent sont probablement dégradées.

Pourtant, dans les opérons *rrn*, on trouve une autre variation. Chez certains, la dernière séquence codée dans le précurseur est l'ARN 5S. Chez d'autres, on peut trouver quelques séquences d'ARNt supplémentaires. Par exemple, dans le *rrnC* il existe deux gènes, l'ARNt^{Trp} et l'ARNt^{Asp}, situés à l'extrémité 3'. (La présence du gène d'ARNt^{Trp} rend cet opéron *rrn* particulier, indispensable à *E. coli*, puisque c'est là la seule copie de ce gène ; il fournit ainsi la seule possibilité d'utilisation du tryptophane dans la synthèse protéique.)

L'enzyme responsable de la maturation des ARNr est l'ARNase III. Dans les cellules *rnc*⁻, les ARNr matures ne se forment pas ; à la place, un précurseur de 30S contenant les séquences 16S, 22S et 5S s'accumule. Ce

précurseur peut être coupé *in vitro* par l'ARNase III et engendrer des molécules servant de précurseurs individuels aux espèces d'ARN matures.

Les précurseurs des ARNr 16S et 22S sont appelés p16 et p22. Chaque précurseur est légèrement plus long que l'ARNr trouvé dans le ribosome, parce qu'il conserve des séquences supplémentaires aux deux extrémités 5' et 3'.

Un mécanisme commun engendre les extrémités des ARN p16 et p22. Dans chaque cas, les régions du transcrit primaire qui renferment les extrémités sont complémentaires. Elles s'apparient pour former une structure en double brin renfermant les régions 5' et 3' du produit. Cette réaction possède une caractéristique singulière : la distance séparant les séquences complémentaires, environ 1 600 nucléotides pour le p16 et 2 900 nucléotides pour le p22.

La Figure 22.7 nous montre que dans chaque cas l'ARNase III coupe des deux côtés de la tige double brin, engendrant simultanément les extrémités 5' et 3' des molécules p16 ou p22. Cette action de l'ARNase III diffère en deux points de celle plus couramment observée dans l'ARN précoce du phage T7. La séquence cible est une tige continue sans aucune « bulle » proéminente. Deux coupures se produisent plutôt qu'une (ce qui arrive également dans un site atypique du T7).

On ne trouve pas d'homologie de séquence étendue, que ce soit entre les sites de l'ARNase III du p16 et du p22, ou bien entre l'un de ces derniers et un site du T7. Les conditions requises par l'ARNase III pour une coupure spécifique de l'ARN sont mystérieuses, mise à part l'exigence commune d'une structure en double brin

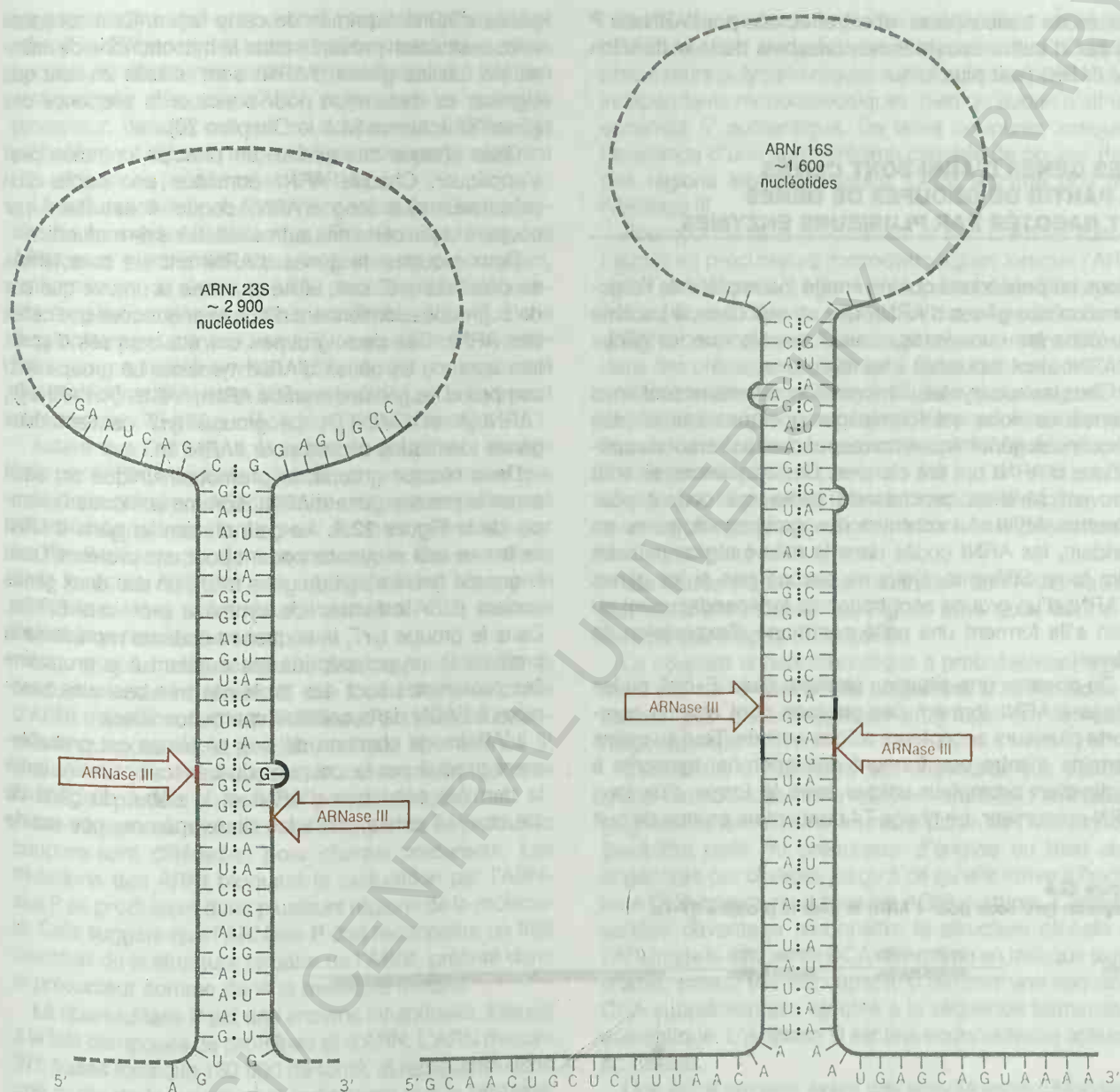


Figure 22.7
Les régions contenant les extrémités 5' et 3' des ARN p16 et p22 s'apparient pour former des tiges en double brin, dont chacune est coupée en deux points par l'ARNase III.

entourant, ou encore comprenant, le site réel de coupure. L'enzyme pourrait dépendre d'une combinaison entre une séquence primaire et une structure secondaire.

Comme les coupures accomplies par l'ARNase III ne libèrent pas de vraies extrémités 5' ou 3' des ARNr 16S

ou 22S, des réactions supplémentaires sont nécessaires à la maturation. Nous ne savons pas grand-chose au sujet de ces réactions supplémentaires si ce n'est qu'elles dépendent de l'action antérieure de l'ARNase III. La maturation des séquences d'ARNt incluses dans les

unités de transcription *rrn* est effectuée par l'ARNase P et par d'autres enzymes responsables de la maturation de l'ARNt (voir plus loin).

LES GÈNES D'ARNt SONT CLIVÉS À PARTIR DE GROUPES DE GÈNES ET RABOTÉS PAR PLUSIEURS ENZYMES

Nous ne possédons qu'une image incomplète de l'organisation des gènes d'ARNt, que ce soit dans la bactérie ou dans les eucaryotes, mais il semble que les gènes d'ARNt aient tendance à se regrouper.

Chez les eucaryotes, l'information sommaire dont nous disposons nous est fournie par quelques cas où des séquences génomiques correspondant à une sonde spécifique d'ARNt ont été clonées. Ces séquences se sont souvent révélées proches de celles qui codent pour d'autres ARNt. Au contraire des groupes de gènes en tandem, les ARNt codés dans la même région peuvent être tous différents. Nous ne savons pas si les gènes d'ARNt d'un groupe sont transcrits indépendamment ou bien s'ils forment une unité commune d'expression de gènes.

On observe une situation similaire dans *E. coli*, où les gènes d'ARNt forment des groupes dont chacun comporte plusieurs accepteurs acides aminés. Tout au moins certains d'entre eux forment des opérons, transcrits à partir d'un promoteur unique sous la forme d'un long ARN précurseur. Le phage T4 contient un groupe de huit

gènes d'ARNt exprimés de cette façon. Cela contraste avec la situation présente dans la mitochondrie de mammifère, où les gènes d'ARNt sont utilisés en tant que signaux de maturation pour ponctuer la séquence des gènes structuraux (voir le Chapitre 20).

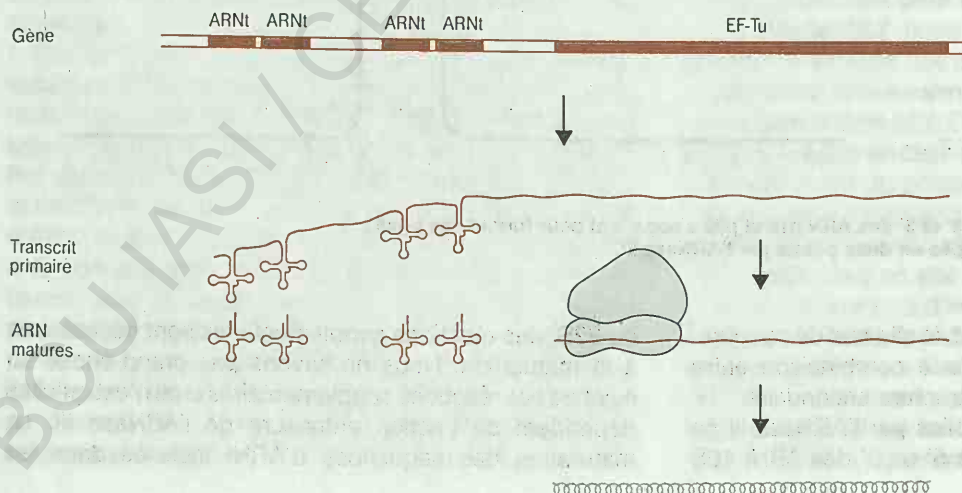
Dans chaque cas connu, un principe commun peut s'appliquer. Chaque ARNt constitue une partie d'un précurseur plus long d'ARN, duquel il est libéré par coupure et par certaines autres réactions de maturation.

Deux groupes de gènes d'ARNt ont été caractérisés en détail dans *E. coli*, et nous avons la preuve que ces deux groupes contiennent d'autres séquences que celles des ARNt. Ces deux groupes ont été baptisés d'après leur contenu en gènes d'ARNt tyrosine. Le groupe *tyrU* comprend les gènes de quatre ARNt, l'ARNt $^{\text{I}}_{\text{Tyr}}$, l'ARNt $^{\text{II}}_{\text{Tyr}}$, l'ARNt $^{\text{III}}_{\text{Tyr}}$ et l'ARNt $^{\text{IV}}_{\text{Tyr}}$. Le groupe *tyrT* contient deux gènes identiques codant pour l'ARNt $^{\text{I}}_{\text{Tyr}}$.

Dans chaque groupe, le promoteur unique est situé avant le premier gène d'ARNt, comme le montre l'exemple de la Figure 22.8. Au-delà du dernier gène d'ARNt se trouve une séquence codant pour une protéine. Dans le groupe *tyrU* il s'agit du gène *tufB* (l'un des deux gènes codant pour le facteur de synthèse protéique EF-Tu). Dans le groupe *tyrT*, la séquence codante représente la protéine P, un polypeptide ressemblant à la protamine (les protamines sont des protéines très basiques associées à l'ADN dans certains spermatozoïdes).

L'ARNm de chacune de ces protéines est probablement produit par la coupure du transcrit primaire entre la dernière séquence d'ARNt et le début du gène de structure. Il existe, dans les deux opérons, une courte

Figure 22.8
L'opéron *tyrU* code pour 4 ARNt et pour la protéine EF-Tu.



séquence commune qui pourrait bien être utilisée à cet effet. A la différence des autres opérons bactériens, dans lesquels les ribosomes se fixent probablement à l'ARNm pratiquement dès que l'ARN polymérase a quitté le promoteur, dans le cas qui nous intéresse, l'ARNm est probablement d'abord engendré par coupure le séparant des séquences précédentes d'ARNt.

La maturation est une étape nécessaire dans la production de tous les ARNt connus. Dans les bactéries, les ARNt sont synthétisés sous la forme de précurseurs, renfermant un ou plusieurs ARNt possédant des séquences supplémentaires aux deux extrémités 5' et 3'. Dans le cytoplasme eucaryote, les ARNt nouvellement synthétisés sédimentent à environ 4,5S, ce qui correspond à des molécules d'environ 100 nucléotides, que l'on peut comparer aux 70 à 80 nucléotides d'un ARNt mature 4S.

Autant que l'on sache, la partie du précurseur constituée par l'ARNt adopte la même structure secondaire que celle présentée par l'ARNt mature ; cela pourrait être une caractéristique importante pour sa reconnaissance par les ribonucléases pendant sa maturation. Des mutations dans la partie ARNt du précurseur peuvent bloquer la maturation, avec pour résultat que les ARNt matures défectueux ne s'accumulent pas.

Une enzyme unique semble être responsable de la production des extrémité 5' de toutes les molécules d'ARNt dans *E. coli*. C'est l'**ARNase P**, une endoribonucléase qui coupe à la jonction, soit de la séquence 5' leader, soit de la séquence intercistronique (lorsque l'ARNt fait partie d'un transcrit polycistronique).

Les séquences primaires dans lesquelles se produit la coupure sont différentes pour chaque précurseur. Les mutations des ARNt bloquant la maturation par l'ARNase P se produisent dans plusieurs régions de la molécule. Cela suggère que l'ARNase P doit reconnaître un trait commun de la structure tertiaire de l'ARNt, présent dans le précurseur comme dans la molécule mature.

La ribonucléase P est une enzyme inhabituelle. Elle est à la fois composée de protéines et d'ARN. L'ARN mesure 375 bases (quelque 130 000 daltons), et représente donc une partie de la masse plus importante que la protéine, qui n'est que d'environ 20 000 daltons. On a trouvé dans les eucaryotes des équivalents potentiels de l'ARNase P.

Dans *E. coli*, deux gènes à l'intérieur desquels des mutations abolissent l'activité de l'ARNase P ont été identifiés : *mnpA* code pour le composant protéique, et *mnpB* code pour l'ARN. Ils sont situés loin l'un de l'autre. Les deux composants, protéine et ARN, de l'ARNase P sont indispensables à son activité catalytique, laquelle semble reposer sur ces bases exceptionnelles, comme nous le verrons au chapitre suivant.

Dans les mutants *mnp* de nombreuses molécules d'ARNt servant de précurseurs s'accumulent. Certains précurseurs polycistroniques sont coupés en précurseurs indépendants monocistroniques, bien qu'aucun n'ait une extrémité 5' authentique. De telles coupures indiquent l'existence d'une autre enzyme capable de couper dans des régions intercistroniques. Cette enzyme doit être l'ARNase III.

Pourquoi certains précurseurs et non d'autres sont-ils coupés en précurseurs monocistroniques lorsque l'ARNase P est absente ? Même si la même série d'enzymes est impliquée dans le clivage de tous les précurseurs d'ARNt, ces enzymes doivent agir selon un ordre différent dans des précurseurs différents. Cette idée est confirmée par l'étude de précurseurs particuliers, dans lesquels des événements analogues peuvent se produire à diverses étapes. L'obligation de respecter une hiérarchie donnée d'événements doit provenir du fait qu'un seul site de clivage est accessible dans le transcrit d'origine, mais que son clivage change alors la structure exposant le site de l'enzyme suivante.

Les autres stades de la maturation de l'ARNt ne sont pas aussi bien établis. La **Figure 22.9** nous donne un résumé des étapes obligatoires.

La coupure endonucléolytique a probablement également lieu sur le côté 3' des séquences d'ARNt. Nous ne savons pas encore quelle(s) enzyme(s) en est (sont) responsable(s).

L'extrémité 3' définitive est engendrée par l'activité d'une exonucléase. Une enzyme appelée l'**ARNase D** grignote les bases, une à une, à partir de l'extrémité 3' (peut-être celle du précurseur d'origine ou bien alors engendrée par clivage), jusqu'à ce qu'elle arrive à l'extrémité CCA commune à tous les ARNt matures. L'enzyme semble davantage reconnaître la structure globale de l'ARNt que la séquence CCA elle-même en tant que signal d'arrêt, puisqu'elle est capable d'éliminer une séquence CCA supplémentaire ajoutée à la séquence terminale 3' authentique. L'ARNase D est une endonucléase agissant au hasard.

Une autre enzyme ayant une activité exonucléolytique est l'**ARNase II**. Jadis considérée comme impliquée dans la maturation de l'ARNt, elle est en fait capable de détruire complètement les séquences d'ARNt. Elle est donc probablement plus en rapport avec la dégradation générale qu'avec une maturation spécifique. Son action est progressive.

Dans la bactérie, on distingue deux types de précurseurs d'ARNt, grâce à leur séquence 3'. Les molécules de type I ont un triplet CCA qui constitue la démarcation entre la séquence d'ARNt mature et le matériel 3' addi-

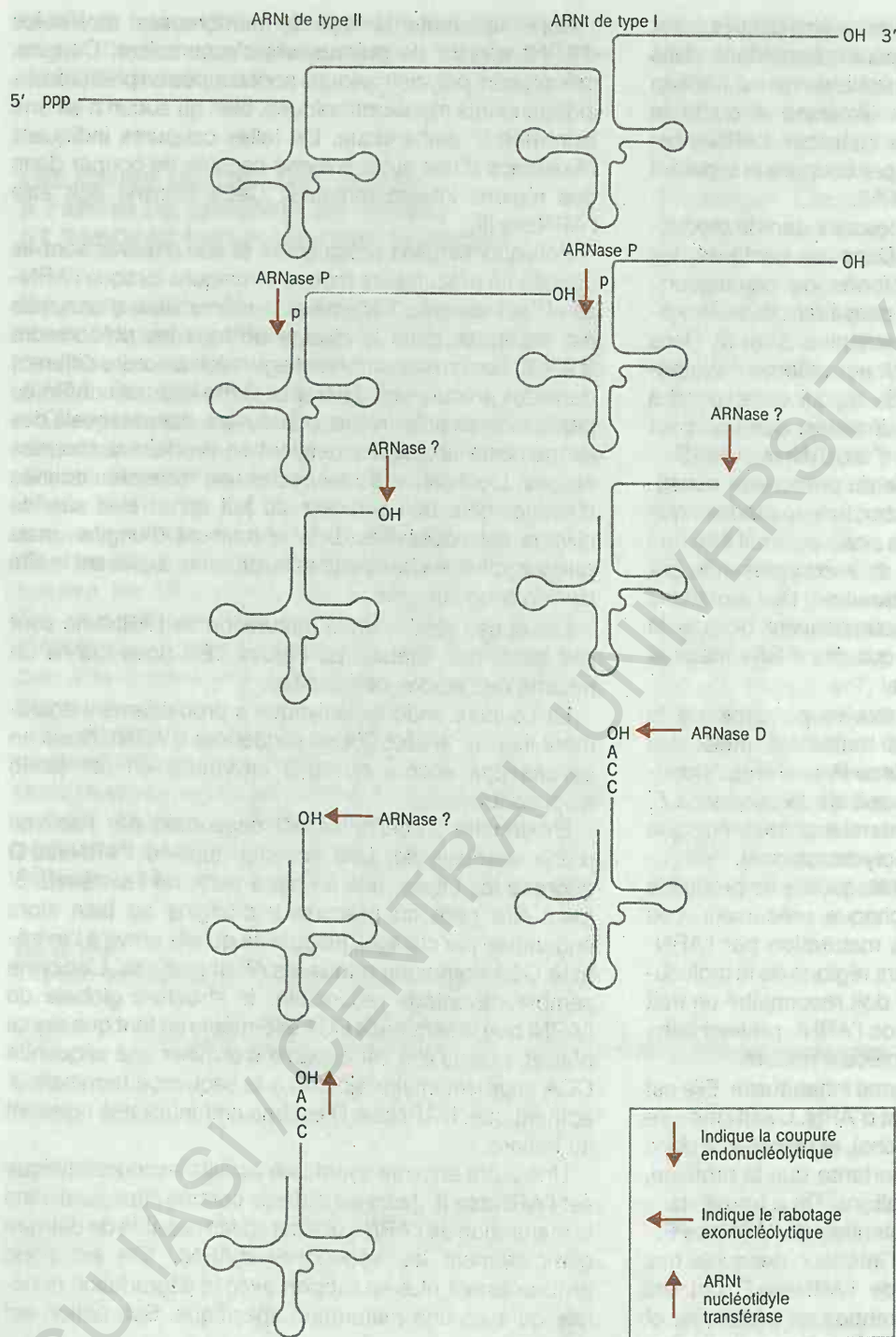


Figure 22.9

Plusieurs enzymes sont nécessaires à la maturation d'un dimère précurseur (hypothétique) contenant un ARNt de type I et II.

tionnel. Les molécules de type II (codées par certains phages) n'ont pas quant à elles de séquence CCA. Après que les nucléotides 3' supplémentaires ont été retirés du précurseur, le triplet CCA doit être ajouté. Nous ignorons totalement de quelle manière est engendrée, dans les ARNt de type II, la terminaison 3' à laquelle sera ajouté le triplet CCA, si c'est une fonction de l'ARNase D comme dans les ARNt de type I, ou bien si une autre enzyme est impliquée. Dans les eucaryotes, probablement tous les ARNt appartiennent au type II.

L'adjonction de la séquence CCA est la fonction d'une enzyme, l'ARNt **nucléotidyl transférase**, qui ajoute le triplet (ou une partie de ce triplet) à toute séquence d'ARNt qui en est dépourvue. Les mutants de *E. coli* ne possédant que de faibles taux de cette enzyme (appelés *cca*⁻) ont une croissance lente, ce qui suggère que cette enzyme est essentielle à la biosynthèse de l'ARNt. Le résidu terminal A de tous les ARNt bactériens est cons-

tamment renouvelé si bien que cette enzyme est non seulement nécessaire à la production initiale des ARNt de type II mais également aux fonctions de réparation.

POUR EN SAVOIR PLUS

Peu de revues ont démêlé les enzymes réellement impliquées dans la maturation de l'ARNt ou de l'ARNr d'après les diverses activités qui ont été détectées. **Gegenheimer & Apirion** ont tenté de faire le tri dans la génétique et la biochimie des enzymes bactériennes (*Microbiol. Rev.* **45**, 502-541, 1981). Une étude sur l'ARNase III qui recouvre également les premières investigations a été réalisée par **Bram, Young & Steitz** (*Cell* **19**, 393-401, 1980). La maturation des ARNt a été passée en revue par **Altman** dans *Transfer RNA* (MIT Press, Cambridge, Mass. 1978, pp. 48-77) et mise à jour dans *Cell* (**22**, 3-4, 1981). Les ribonucléases de *E. coli* ont été récapitulées par **Deutscher** (*Cell* **40**, 731-732, 1985).

CHAPITRE 23

L'ARN EST UN CATALYSEUR : LES MÉCANISMES DE L'ÉPISSAGE

Avoir découvert que l'épissage de l'ARN élimine les introns des transcrits de gènes discontinus soulève une série de questions concernant l'expression des gènes. En terme de mécanisme, nous voudrions savoir en quel point a lieu l'épissage, par rapport aux autres modifications de l'ARN [le « *capping* » (addition de la coiffe ou cap en 5' des ARN) et la polyadénylation]. Les introns sont-ils excisés d'un précurseur dans un ordre particulier ? L'épissage se produit-il en un endroit spécifique du noyau ; a-t-il un rapport avec d'autres événements, par exemple, le transport du noyau vers le cytoplasme ? L'absence d'épissage crée-t-elle une différence importante dans l'expression des gènes continus ? La maturation de l'ARN (par l'épissage ou par d'autres mécanismes) est-elle utilisée pour *contrôler* l'expression du gène en établissant une sélection entre les précurseurs disponibles ?

L'excision-épissage de l'ARN implique la coupure des liaisons phosphodiester aux frontières exon-intron, et la formation d'une liaison entre les extrémités des exons. Cela pourrait résulter soit de deux réactions successives et indépendantes, soit d'une réaction de transfert coordonnée. La rupture des liaisons phosphodiester est commune à d'autres réactions de maturation de l'ARN, mais leur formation dans l'ARN est propre aux processus d'excision-épissage. Quel(s) que soi(en)t le(s) méca-

nisme(s) enzymatique(s), une caractéristique décisive est que les groupements participant à la réaction doivent être juxtaposés pour former des liaisons.

Il n'existe pas un mécanisme unique d'épissage de l'ARN. Nous pouvons en définir plusieurs :

- L'excision des introns des précurseurs de l'ARN nucléaire de levure met en œuvre des activités enzymatiques, dont les relations avec le substrat semblent s'apparenter à celles des enzymes de maturation de l'ARNt, puisque la conformation du précurseur d'ARNt en est une des caractéristiques.
- L'excision de l'intron d'un ARNr précurseur dans des micro-organismes ciliés est une propriété autonome de l'ARN. Une réaction de même ordre supprime les introns de certains ARN mitochondriaux de la levure.
- Les introns sont éliminés des ARN nucléaires des eucaryotes supérieurs par un système qui ne présente apparemment rien de plus spécifique que de courtes séquences consensus préservées aux frontières exons-introns et à proximité.

Ainsi, si tous les gènes mosaïques ont une origine commune, de considérables divergences se sont produites depuis dans la manière de reconstituer une séquence codante intacte d'ARN.

L'EXCISION-ÉPISSAGE DE L'ARNt DE LA LEVURE IMPLIQUE COUPURE ET JONCTION

Environ 40 des quelque 400 gènes nucléaires d'ARNt de la levure sont morcelés. Chaque gène possède un intron unique situé exactement sur un nucléotide par-delà l'extrémité 3' de l'anticodon. La longueur des introns varie de 14 à 46 pb. Ces derniers, dans les gènes d'ARNt du même groupe, ont des séquences proches, mais les introns des gènes d'ARNt chargeant des acides aminés différents n'ont aucun rapport entre eux. *Il n'existe aucune séquence consensus qui puisse être reconnue par les enzymes responsables de l'épissage.*

Tous les introns renferment une séquence complémentaire de l'anticodon de l'ARNt. Cela crée une structure alternative au bras de l'anticodon dans laquelle l'anticodon est apparié pour former une extension du bras habituel. La **Figure 23.1** nous en trace un exemple. Seul le bras de l'anticodon est affecté, le reste de la molécule garde sa structure habituelle.

Il est possible, en introduisant des modifications dans la molécule d'ARN, soit par manipulation *in vitro* du gène cloné, soit par mutation *in vivo*, de déterminer les caractéristiques du précurseur reconnues pour l'épissage. De telles expériences révèlent que la séquence et la taille exacte de l'intron sont sans importance. Les mutations altérant la structure secondaire de la région intronique n'empêchent pas l'épissage. *L'épissage pourrait donc dépendre de la reconnaissance d'une structure secondaire commune plutôt que d'une séquence commune.*

Cela nous ramène aux conclusions que nous avons avancées dans les Chapitres 6 et 22 concernant la

structure et l'expression de l'ARNt. Il ne semble pas que certaines bases spécifiques soient impliquées de manière significative dans les réactions de reconnaissance, qui en fait reposent sur la *conformation globale de la molécule en tant que tout*. En particulier, il peut être important d'identifier les caractéristiques de structure du précurseur, conservées dans l'ARNt mature. Ceci est cohérent avec la constatation d'après laquelle l'ARNt mature opère une inhibition compétitive de l'épissage des précurseurs, probablement parce qu'il est reconnu par le(s) enzyme(s).

L'existence, dans la levure, d'un mutant thermo-sensible incapable d'éliminer les introns et accumulant des précurseurs morcelés dans le noyau, nous offre une possibilité d'étudier la réaction d'épissage. Nous ne connaissons pas la base moléculaire de cette anomalie ; mais elle est propre aux gènes morcelés, puisque les ARNt codés par des gènes continus peuvent subir une maturation et être transportés vers le cytoplasme.

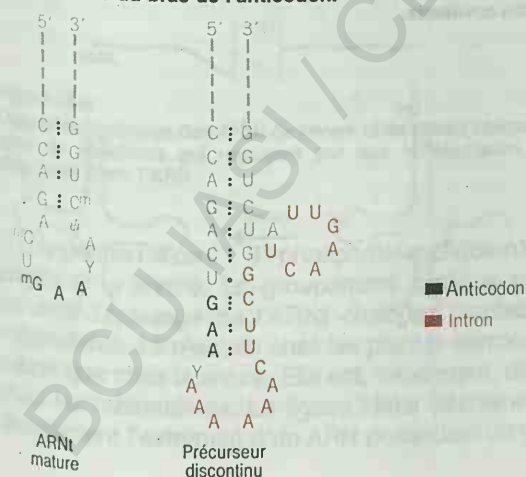
Cette propriété est très utile car elle permet d'isoler les précurseurs de l'ARNt, sous une forme dans laquelle leur seule différence de séquence avec l'ARNt mature est la présence de l'intron. (Certaines modifications de bases peuvent également être absentes des précurseurs.) Il est possible d'utiliser ces molécules comme substrat d'un système acellulaire provenant des cellules du type sauvage. L'épissage du précurseur peut être suivi relativement simplement grâce à la réduction de taille qui en résulte. C'est le changement de position de la bande sur l'électrophorèse qui le concrétise, comme l'illustre la **Figure 23.2**. La réduction de taille de la forme mature d'ARNt s'accompagne de l'apparition d'une nouvelle bande qui représente l'intron.

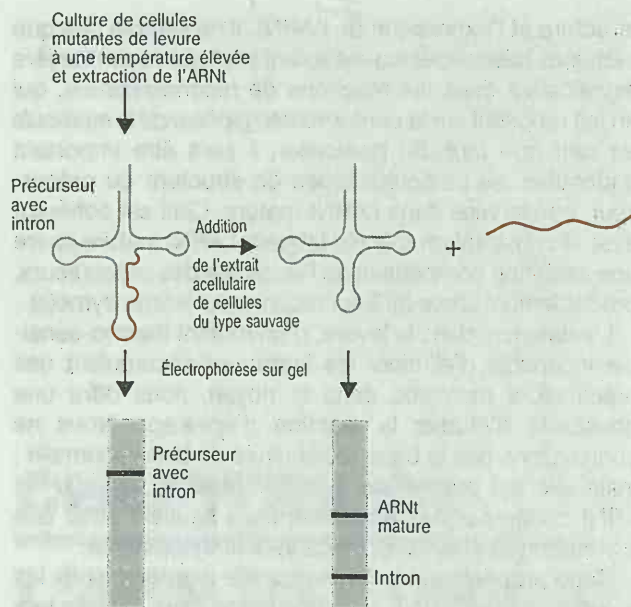
L'extrait acellulaire peut être fractionné en analysant sa capacité à épisser l'ARNt. La réaction *in vitro* nécessite de l'ATP. En caractérisant les réactions qui ont lieu avec ou sans ATP, nous pouvons observer que les *deux phases distinctes de la réaction sont catalysées par des enzymes différentes*.

La première étape ne nécessite pas d'ATP. Elle implique la rupture de la liaison phosphodiester, prenant ainsi la forme d'une réaction nucléasique atypique. Cette réaction est catalysée par une endonucléase qui se comporte comme une protéine liée à la membrane. La seconde étape nécessite de l'ATP et implique la formation d'une liaison ; elle est appelée réaction de **ligation**, et l'activité enzymatique qui en est responsable est définie comme une **ARN ligase**. L'ensemble de ces mécanismes est décrit à la **Figure 23.3**.

En l'absence d'ATP, l'endonucléase coupe le précurseur aux deux extrémités de l'intron, engendrant deux demi-molécules d'ARNt et une séquence intercalaire

Figure 23.1
L'intron de l'ARNt^{Phe} de levure s'apparie à l'anticodon pour modifier la structure du bras de l'anticodon.



**Figure 23.2**

Nous pouvons suivre l'épissage *in vitro* de l'ARNt de levure en analysant le précurseur et les produits de l'ARN par une électrophorèse sur gel.

linéaire. Ces intermédiaires ont des extrémités uniques. Chaque extrémité 5' s'achève par un groupe hydroxyle ; et chaque extrémité 3' par un groupe de phosphate cyclique 2', 3'. (Toutes les autres enzymes connues impliquées dans la maturation de l'ARN coupent sur l'autre côté de la liaison phosphate, comme il est mentionné au Chapitre 22.)

Lorsque l'on ajoute l'ATP, la seconde réaction se produit. Les deux moitiés d'ARNt peuvent s'apparier pour former une structure comparable à celle de l'ARNt. L'action de l'ARN ligase réunit de façon covalente les deux moitiés en créant une liaison phosphodiester.

Les réactions impliquées dans chaque étape du mécanisme de l'épissage sont illustrées par la **Figure 23.4**. Les deux extrémités atypiques engendrées par l'endonucléase doivent être détruites.

Le groupement phosphate cyclique est ouvert pour créer une terminaison 2'-phosphate. Cette réaction peut nécessiter une phosphodiesterase cyclique. Le produit possède un groupe 2'-phosphate et un groupe 3'-OH.

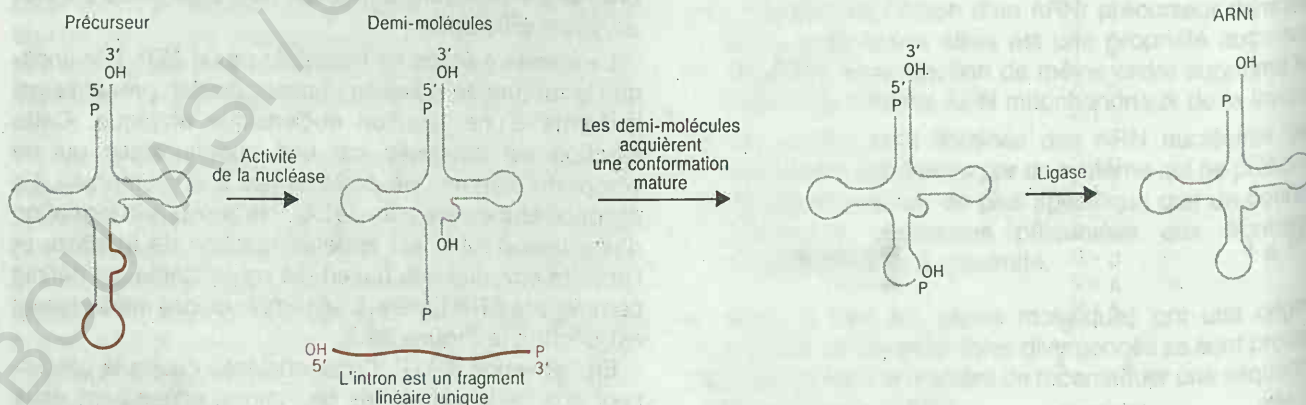
Le groupe 5'-OH engendré par la nucléase doit être converti en un 5'-phosphate. La réaction a probablement lieu grâce à une réaction kinasique classique, dans laquelle le phosphate- γ de l'ATP est le donneur.

La ligase transfère un AMP sur le groupe 5'-phosphate, formant une liaison 5'-5'-phosphate-phosphate. Ensuite l'AMP est déplacé par une attaque du groupement 3'-OH de l'autre moitié de l'ARNt. Notons que le groupement phosphate reliant les deux exons n'est pas un constituant originel de l'ARN, mais a été produit par l'ATP pendant la réaction.

La molécule épissée est à présent continue, elle possède une liaison 5'-3'-phosphate au site de l'épissage, mais elle possède également un groupe 2'-phosphate marquant l'événement. Le groupe excédentaire doit être éliminé par une phosphatase. (Pendant la période intermédiaire il peut servir en marquant le site où s'est produite la ligation.)

Figure 23.3

L'excision-épissage se produit grâce à l'action d'une nucléase qui produit l'intron et les demi-molécules d'ARNt, suivie de l'action d'une ligase qui réunit les demi-molécules de façon covalente.



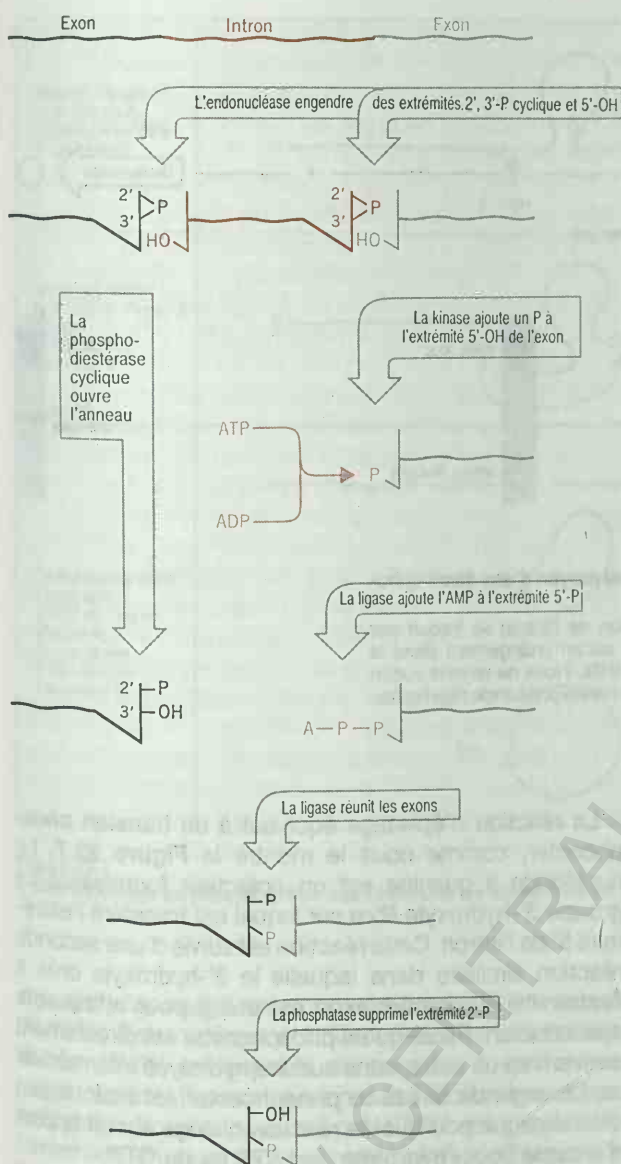


Figure 23.4
L'excision-épissage des ARNt de levure et de plante comprend une série de réactions qui reposent sur des terminaisons 5' et 3' atypiques dans l'ARN.

L'apparition d'un 2', 3'-phosphate cyclique n'est pas propre à la levure. Le groupement cyclique est créé pendant l'épissage de l'ARNt chez les plantes et les mammifères. La réaction chez les plantes semble être la même que chez la levure. Elle est, cependant, différente chez les mammifères. La ligase HeLa (humaine) réunit directement l'extrémité d'un ARN possédant un groupe-

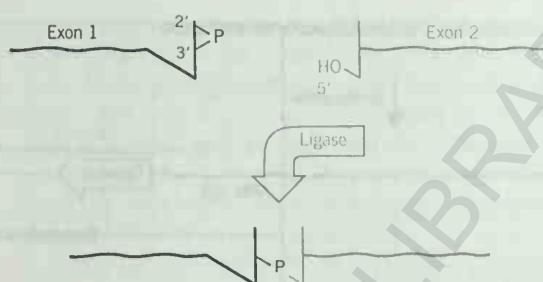


Figure 23.5

La réaction d'épissage de l'ARNt chez les mammifères pourrait impliquer une réaction directe entre le 2', 3'-phosphate cyclique et un groupe 5'-hydroxyle.

ment 2', 3'-phosphate cyclique à l'extrémité d'un ARN possédant un groupement 5'-hydroxyle. Ainsi, le groupement phosphate réunissant les deux exons est le phosphate qui était, à l'origine, à l'extrémité de l'exon gauche, comme l'illustre la **Figure 23.5**.

Il est également possible d'obtenir l'épissage des précurseurs de l'ARNt de levure dans un extrait obtenu à partir de la vésicule germinative (noyau) des oocytes de *Xénopus*. Cela montre que la réaction n'est pas spécifique d'une espèce. *Xénopus* possède probablement les enzymes capables de reconnaître les introns des ARNt de levure.

L'ARN PEUT POSSÉDER UNE ACTIVITÉ CATALYTIQUE

En biochimie, l'idée que seules les protéines ont une activité enzymatique est profondément enracinée. (Cependant, nous pourrions nous rappeler que les passionnés de la fonction de la protéine pensèrent à un moment que seules les protéines pouvaient avoir la liberté de constituer le matériel génétique !) Une des raisons d'être de l'assimilation des enzymes repose sur le fait que seules les protéines avec leurs structures tridimensionnelles variées et leurs différents groupements latéraux, ont la souplesse nécessaire pour créer les sites actifs requis pour catalyser des réactions biochimiques. Cependant, la caractérisation de deux systèmes impliqués dans la maturation de l'ARN a révélé que cette vision des choses était une simplification hâtive.

Les gènes des deux principaux ARNr du *Tetrahymena thermophila* suivent l'organisation habituelle selon laquelle tous deux sont exprimés en tant que partie d'une unité de transcription commune. Le produit est un ARN précurseur de 35S possédant la séquence du petit ARNr dans la partie 5', et la séquence du plus grand ARNr (26S) vers l'extrémité 3'.

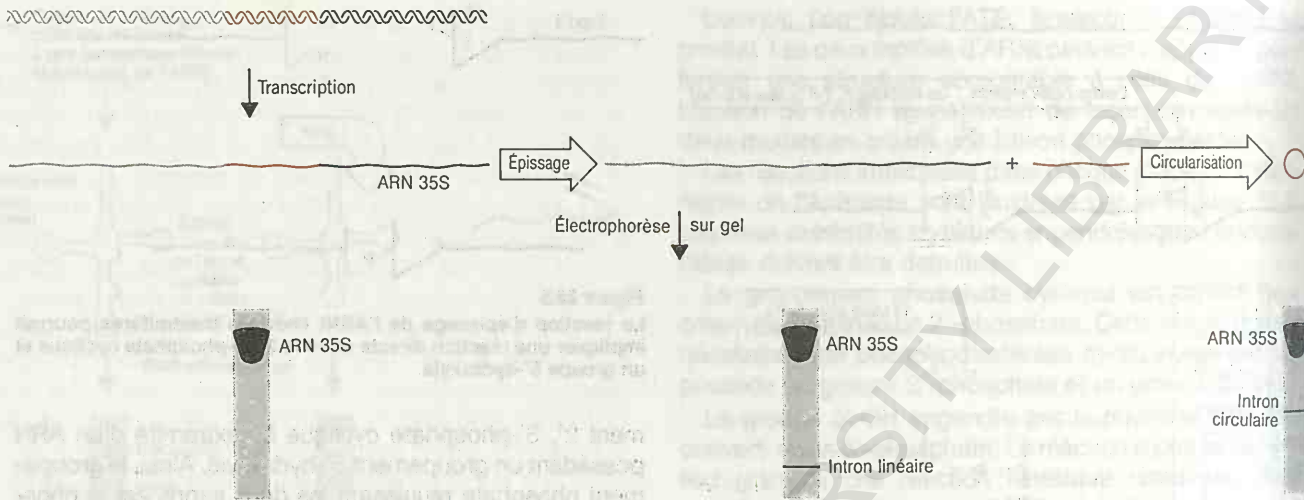


Figure 23.6

L'épissage du précurseur de l'ARNr 35S de *Tetrahymena* peut être suivi au moyen d'une électrophorèse sur gel.

Le précurseur de l'ARN 35S forme une bande relativement large. L'excision de l'intron se traduit par l'apparition d'une petite bande migrant rapidement. (Nous ne distinguons aucun changement dans la bande de l'ARN 35S à cause de son épaisseur et de sa faible diminution de taille. Nous ne voyons aucun exon libre.) Lorsque l'intron devient circulaire, il migre plus lentement, comme le révèle une bande plus haute.

Dans certaines souches de *T. thermophila*, la séquence codant pour l'ARNr 26S est interrompue par un court intron unique. Lorsque l'ARN précurseur 35S est incubé *in vitro*, l'épissage se déroule comme une réaction autonome. L'intron est excisé du précurseur et s'accumule en un fragment linéaire de 400 bases, qui est ensuite converti en un ARN circulaire. La **Figure 23.6** résume ces phénomènes.

Cette réaction ne nécessite qu'un cation monovalent, un cation divalent, et un nucléotide à guanine. Aucune autre base ne peut se substituer au G ; mais il n'est pas nécessaire d'avoir un triphosphate ; GTP, GDP, GMP, et la guanosine elle-même peuvent tous être utilisés. Cela signifie que, bien qu'un résidu G soit impliqué, il n'y a pas de demande précise d'énergie. La guanine doit posséder un groupe 3'-hydroxyle.

Le destin de la guanine peut être suivi grâce à un marquage radioactif. La radioactivité pénètre initialement le fragment de l'intron linéaire excisé. Le résidu G se fixe à l'extrémité 5' de l'intron par une liaison phosphodiester normale. Lorsque l'intron linéaire devient circulaire, l'extrémité 3' se rattache à un site situé à 15 bases de l'extrémité 5', et un segment terminal de 15 bases, comprenant le G supplémentaire, est libéré.

La réaction d'épissage équivaut à un transfert phosphoester, comme nous le montre la **Figure 23.7**. Le nucléotide à guanine est un cofacteur fournissant un groupe 3'-hydroxyle libre sur lequel est transféré l'extrémité 5' de l'intron. Cette réaction est suivie d'une seconde réaction similaire dans laquelle le 3'-hydroxyle créé à l'extrémité du premier exon est utilisé pour attaquer le second exon. Parce qu'un phosphoester est directement converti en un autre, sans aucune hydrolyse intermédiaire, l'énergie de la liaison phosphoester est maintenue ; cela explique pourquoi la réaction n'exige aucun apport d'énergie pour l'hydrolyse de l'ATP ou du GTP.

Les deux transferts semblent être liés. Nous n'avons observé aucun exon libre ; leur ligation doit, par conséquent, appartenir à la même réaction que celle libérant l'intron. La circularisation du fragment de l'intron linéaire pourrait s'expliquer par un autre transfert de phosphoester, au cours duquel la liaison située à 15 bases de l'extrémité est rompue, et l'extrémité 5' libérée, transférée sur le groupe 3'-hydroxyle de la molécule de l'intron. (Il existe un parallèle entre le transfert du groupe 3'-OH sur le 5'-P sans apport net d'énergie, et les enzymes coupant et réparant l'ADN dont il est question au Chapitre 28.)

Le système *in vitro* ne comprend aucune protéine, la

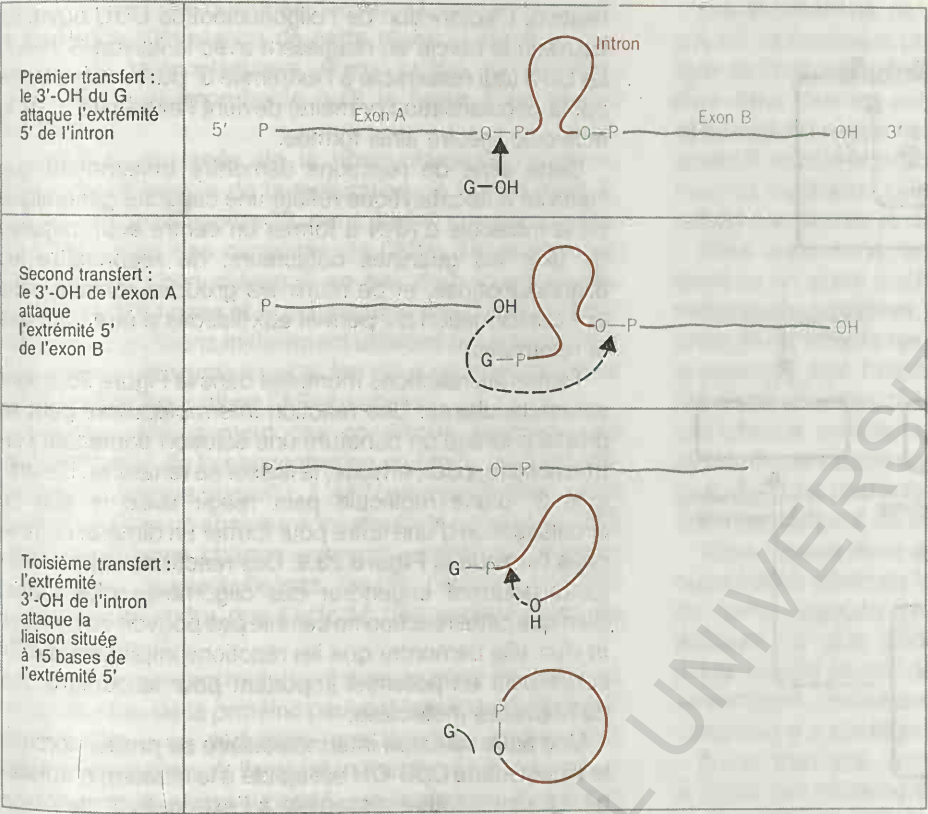


Figure 23.7
L'auto-épissage se produit grâce aux réactions de transfert d'un phosphoester.

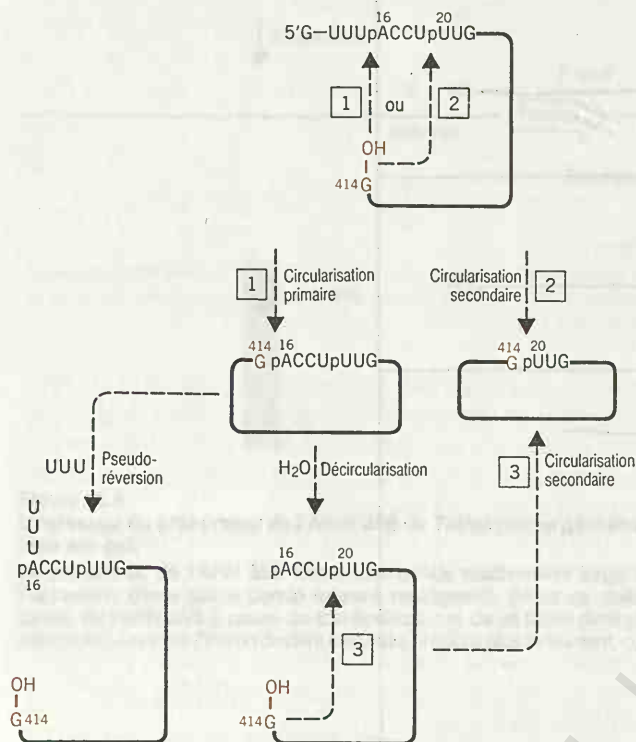
capacité d'épissage est donc une propriété intrinsèque de l'ARN. Cette action a été dénommée une **autocatalyse**, car l'ARN est capable de subvenir à son propre réagencement. Il ne catalyse pas la réaction des autres molécules. Nous supposons que l'ARN est capable de former une structure secondaire/tertiaire spécifique, dans laquelle les groupes concernés sont mis en juxtaposition, afin qu'une guanine soit liée à un site spécifique, et que par la suite, les réactions provoquant la rupture de la liaison, puis la réunion montrées à la Figure 23.7 puissent avoir lieu. Bien qu'elle soit propre à l'ARN, cette réaction pourrait être favorisée *in vivo* par des protéines (qui, par exemple, pourraient stabiliser la structure de l'ARN).

La capacité de s'engager dans des réactions de transfert de phosphoesters réside dans la séquence de l'intron, qui continue à être réactif même après son excision, sous forme de molécule linéaire. La Figure 23.8 résume ses activités.

L'intron peut devenir circulaire à la suite du transfert d'un phosphoester supplémentaire. Le G de l'extrémité 3' peut attaquer indifféremment les deux positions situées à proximité de l'extrémité 5'. La réaction la plus fréquente est la *circularisation primaire* entre le G⁴¹⁴ terminal et le A¹⁶. Une *circularisation secondaire*, moins fréquente, implique la réaction du G⁴¹⁴ avec l'U²⁰, qui est la base située à quatre positions de l'extrémité 5' de l'intron linéaire.

Le cercle peut engendrer de nouveau, *in vitro*, une molécule linéaire, par hydrolyse spécifique de la liaison (G⁴¹⁴-A¹⁶) qui avait refermé le cercle. Ainsi, un seul phosphate sur plus de 400 de la chaîne polynucléotidique est spécifiquement réactif. Cette molécule linéaire plus courte garde encore la capacité de réaliser une circularisation secondaire.

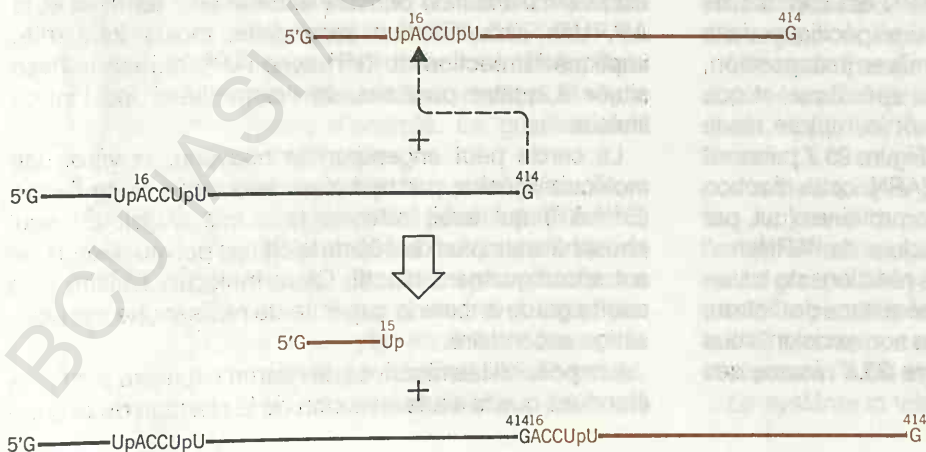
Les possibilités réactives de l'intron circulaire sont plus étendues que la seule inversion de la réaction de circularisation.

**Figure 23.8**

L'intron excisé peut former des cercles en utilisant l'un des deux sites internes pour réagir avec l'extrémité 5', et peut ouvrir à nouveau les cercles en réagissant avec de l'eau ou des oligonucléotides.

Figure 23.9

Une réaction entre deux introns linéaires peut engendrer des dimères.



risation. L'adjonction de l'oligonucléotide UUU ouvre de nouveau le cercle en réagissant avec la liaison $G^{414}-A^{16}$. Le UUU (qui ressemble à l'extrémité 3' du 15-mère libéré par la circularisation primaire) devient l'extrémité 5' de la molécule linéaire ainsi formée.

Cette série de réactions démontre brillamment que l'activité autocatalytique reflète une capacité généralisée de la molécule d'ARN à former un centre actif, capable de fixer les guanines cofacteurs, de reconnaître les oligonucléotides, et de réunir les groupes réactifs dans une conformation qui permet aux liaisons d'être rompues et reformées.

Toutes les réactions montrées dans la Figure 23.8 sont intramoléculaires. Une réaction *intermoléculaire* peut se produire lorsqu'on dénature une solution contenant l'intron linéaire, pour, ensuite, la laisser se renaturer. L'extrémité 3' d'une molécule peut réagir avec le site de circularisation d'une autre pour former un dimère, comme nous l'indique la Figure 23.9. Des réactions supplémentaires pourront engendrer des oligomères plus longs. Bien que cette réaction ne semble pas pouvoir se dérouler *in vivo*, elle démontre que les réactions impliquant l'ARN constituent un potentiel important pour la construction de nouvelles molécules.

Une autre réaction intermoléculaire se produit lorsque le dinucléotide CpU-OH est ajouté à la réaction d'épissage. Le dinucléotide ressemble à l'extrémité 3' de l'exon A (ACCU). Il peut essentiellement se substituer à l'exon A au cours de la seconde réaction de transfert de la Figure 23.7. Le produit en est la séquence linéaire CU-exon B. Cela suggère que la réaction de ligation ne requiert qu'un

site composé des quelques dernières bases de l'exon A. En principe, l'existence de cette réaction implique que l'épissage intermoléculaire (*trans*) puisse se produire, bien qu'il soit improbable qu'il le fasse à un niveau significatif.

L'analyse poussée de la ribonucléase P, l'endonucléase responsable de la maturation de l'ARNt dans *E. coli* décrite au Chapitre 22, nous fournit une démonstration frappante des capacités de l'ARN. Nous pouvons dissocier la ribonucléase P en ses deux composants, l'ARN de 375 bases et le polypeptide de 20 000 daltons. Dans les conditions initialement utilisées pour déterminer l'activité de l'enzyme *in vitro*, les deux composants sont nécessaires pour cliver l'ARNt substrat.

Mais un changement des conditions ioniques, une augmentation de la concentration en Mg^{2+} , rend le composant protéique superflu. *L'ARN seul peut catalyser la réaction !* Si nous analysons les résultats comme si l'ARN était une enzyme, chaque « enzyme » catalyse la coupure d'au moins quatre substrats. En fait, l'activité de l'ARN n'est pas moindre que l'activité des extraits bruts de ribonucléase P.

Sachant que des mutations dans le gène de l'ARN ou dans le gène de la protéine peuvent inactiver l'ARNase P *in vivo*, nous en déduisons que les deux composants sont nécessaires à l'activité naturelle de l'enzyme. A l'origine nous avions supposé que la protéine fournissait l'activité catalytique, tandis que l'ARN remplissait un rôle annexe, par exemple, en aidant à la fixation du substrat (il possède en effet quelques courtes séquences, complémentaires des régions exposées de l'ARNt). A présent, ces rôles se révèlent inversés !

Comment l'ARN peut-il fournir un centre catalytique ? Cela semble raisonnablement possible si nous considérons un centre actif comme une surface exposant une série de groupes actifs dans un rapport fixe. Dans une protéine, les groupes actifs sont constitués par les chaînes latérales des acides aminés, qui présentent une notable variabilité, incluant des groupes ioniques positifs et négatifs et des groupes hydrophobes (voir la Figure 1.2). Dans un ARN, les motifs disponibles sont plus restreints, ils sont constitués essentiellement par des groupes de bases exposés. Nous pourrions supposer que de courtes régions sont maintenues dans une structure spécifique par la conformation secondaire/tertiaire de la molécule, fournissant une surface de groupes actifs capables de maintenir un environnement dans lequel les liaisons peuvent être rompues et recrées dans une autre molécule. Il semble inévitable que l'interaction entre l'ARN catalyseur et l'ARN substrat dépende de l'appariement des bases qui constitueront l'environnement.

Les implications de ces découvertes sur l'évolution posent de nombreux problèmes. La « personnalité » multiple de l'appareil génétique, dans lequel l'ARN est présent dans tous les composants, mais où les protéines assument les réactions catalytiques, a toujours été étonnante. Il semble improbable que les tous premiers systèmes de réplication puissent avoir contenu à la fois des acides nucléiques et des protéines.

Mais supposons que les premiers systèmes aient contenu un acide nucléique autorépliquatif, aux activités catalytiques primitives, exclusivement celles capables de créer et de rompre les liaisons phosphodiester. Si nous supposons que l'implication des liaisons 2' dans les réactions actuelles d'épissage provient de ces activités catalytiques primitives, nous pouvons alors faire l'hypothèse que l'acide nucléique originel était l'ARN, puisque l'ADN ne possède pas de groupe 2'-OH et ne peut donc pas assurer de telles réactions.

Nous voyons donc que les protéines pourraient s'être rajoutées en raison de leur aptitude à stabiliser la structure de l'ARN, jusqu'ici probablement maintenue de façon précaire. La plus grande souplesse des protéines leur aurait ensuite permis de prendre en charge des réactions catalytiques, conduisant ainsi en fin de compte au système complexe et sophistiqué de l'expression moderne du gène.

Nous trouvons encore des restes de ce système originel, plus clairement dans les exemples de l'ARNase P et de l'ARNr de *Tetrahymena*, et, de manière plus spéculative, dans des organites tels que le ribosome. Nous pouvons considérer le ribosome comme constitué à l'origine d'un ARN catalytique, avec des protéines s'ajoutant lentement à la structure, assurant les fonctions catalytiques, et reléguant en fin de compte l'ARN à un rôle de structure — ce serait une inversion exacte des rôles originels des deux composants.

CERTAINS INTRONS MITOCHONDRIAUX PRÉSENTENT DES SIMILITUDES AVEC LES INTRONS CAPABLES D'AUTO-ÉPISSAGE

Les introns présents dans les gènes mitochondriaux fongiques ont été divisés en deux catégories, appelées classe I et classe II, selon leur organisation interne. Les introns de chacun de ces groupes peuvent être organisés en une structure secondaire distincte et renfermer certaines courtes séquences consensus caractéristiques. Les longueurs des introns dans chaque groupe varient énormément, et les séquences consensus peuvent être situées à une distance considérable des jonctions réelles d'épissage.

Les introns de la classe I (qui sont la majorité) ne

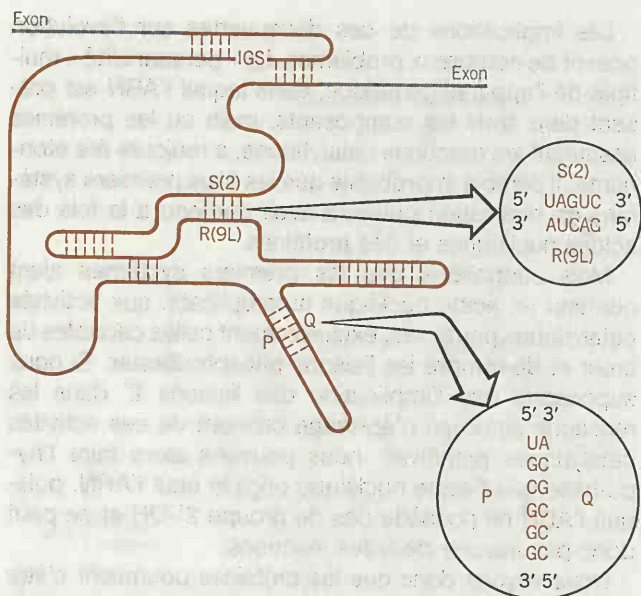


Figure 23.10.

Les introns de la classe I peuvent posséder une structure secondaire commune, dans laquelle l'appariement entre les paires des séquences consensus P-Q et S-R est nécessaire à l'épissage. Une séquence interne guide (IGS)* peut mettre en contact les extrémités de l'intron.

* IGS : Internal Guide Sequence.

montrent aucune séquence conservée aux jonctions d'épissage, mais renferment de courts segments internes de séquence conservée. Ces éléments ont des contreparties dans l'intron nucléaire auto-épissé de *T. thermophila* et dans l'intron de l'ARNr nucléaire d'un autre champignon, le *Physarum polycephalum*.

La Figure 23.10 nous montre un modèle de la structure secondaire de l'intron de *Tetrahymena*, dans lequel les séquences consensus interagissent pour former des régions appariées. Il est possible que la structure secon-

daire rapproche, les unes des autres, les extrémités des introns pour rendre l'épissage possible. Cela pourrait se réaliser si l'extrémité de chaque intron s'appariait avec une partie d'une « séquence interne guide ».

Les éléments P et Q mesurent chacun 10 bases ; 6 à 7 bases sont impliquées dans l'appariement. Les éléments R et S mesurent chacun 12 bases, mais seulement 5 bases sont impliquées dans l'appariement ; ces éléments sont également appelés 9L et 2, tirant leurs noms des mutations mitochondriales (voir plus loin). Les séquences réelles trouvées dans les trois types d'intron possédant ces éléments conservés sont comparées dans le Tableau 23.1.

Nous savons que certaines séquences consensus sont nécessaires à l'épissage, que ce soit directement ou indirectement, parce qu'elles fournissent les sites de mutations agissant en *cis* qui bloquent l'épissage (voir plus loin).

Le rapport entre les introns nucléaires de *Tetrahymena* et du *Physarum* et les introns mitochondriaux de classe I suggère que les deux groupes ont une origine évolutive commune. Une migration a dû se produire entre le noyau et la mitochondrie, bien que nous n'ayons aucun moyen de savoir dans quelle direction elle s'est produite. La dispersion de l'intron nous fait supposer, qu'au moins dans certains cas, il était inséré dans un gène préexistant.

L'origine de la séquence peut avoir été un intron, qui, grâce à son aptitude à être excisé de l'ARN, a engendré une forme libre capable de s'insérer en d'autres sites. Une autre alternative serait que cette séquence ait été, à l'origine, une séquence indépendante (par exemple un petit cercle) possédant l'aptitude de s'auto-exciser ; après son insertion, la séquence se serait fixée sous sa forme actuelle.

Entre l'intron de *Tetrahymena* et les introns mitochondriaux de la classe I, le mécanisme d'épissage a-t-il été conservé ? La réaction globale est similaire. Dans l'exem-

Tableau 23.1.

On trouve des éléments conservés dans les introns mitochondriaux de classe I, dans l'intron de l'ARNr de *Tetrahymena* et dans la thymidylate synthétase du phage T4.

	P	Q	R(9L)	S(2)
Classe I mt <i>Tetrahymena</i> Phage T4	UGCUGG UGC GGG UGC UAA	UCAGCAG CCACGCA GAGGCAU	GACUA GACUA GACUA	UAGUC UAGUC UAGUC

Toutes les séquences sont écrites de 5' en 3' ; des exemples de réactions d'appariement (avec un membre de chaque paire en orientation inverse) sont montrés à la Figure 23.10.

ple du grand ARNr de la mitochondrie de levure, on trouve l'intron de l'ARN, excisé sous deux formes exactement analogues aux structures de l'intron de *Tetrahymena* décrites dans les Figure 23.7 et 23.8. Une forme linéaire possède un G terminal en 5' qui n'est pas codé dans l'ADN. De plus, une forme circulaire a perdu les trois nucléotides terminaux en 5' de l'intron.

Les introns mitochondriaux de la classe I deviennent-ils capables d'auto-excision *in vitro* ? L'intron *cob1* de *Neurospora crassa* s'auto-excise au cours d'une réaction qui requiert une guanine, incorporée dans l'extrémité 5' de l'intron linéaire excisé. Dans le grand ARNr mitochondrial de *S. cerevisiae* la réaction est similaire ; et l'intron linéaire excisé se circularise ensuite.

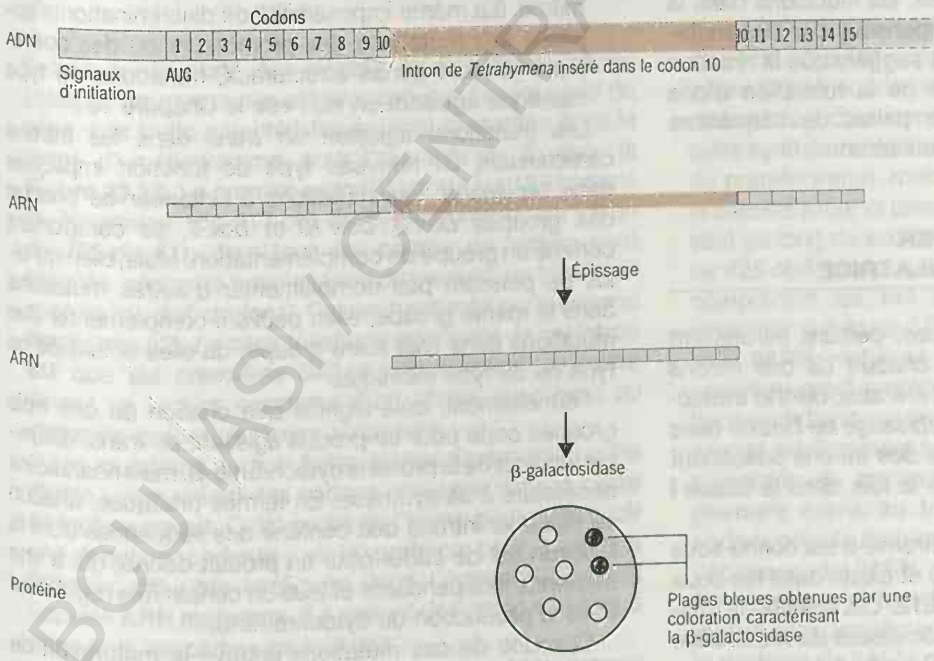
Ces similitudes suggèrent que tous les introns de ce type partagent un mécanisme commun d'épissage, par transfert d'un phosphoester (comme la Figure 23.7 nous en donnait l'exemple) dans lequel la structure de l'ARN joue un rôle déterminant. Cependant, nous savons que plusieurs introns mitochondriaux de la classe I sont sous la dépendance de fonctions agissant en *trans*, des

mutations dans les gènes extérieurs empêchant l'épissage *in vivo*. Cette réaction n'est probablement pas, *in vivo*, une réaction d'auto-épissage, mais elle requiert des fonctions enzymatiques ou protéiques. La capacité d'auto-épissage *in vitro* peut représenter une forme résiduelle des interactions biochimiques primordiales ; comme nous l'avons suggéré pour la ribonucléase P, la structure de l'ARN crée peut-être le site actif, mais ne peut le faire *in vivo* que si elle est aidée par un complexe protéique.

L'importance de l'appariement des bases dans la création du noyau vital de l'ARN est soulignée par les propriétés des mutations agissant en *cis* qui inhibent l'épissage des introns de la classe I. De telles mutations ont été isolées pour les introns mitochondriaux grâce à des mutants qui ne peuvent pas éliminer un intron *in vivo*, et, dans l'intron de *Tetrahymena*, en transférant la réaction d'épissage dans l'environnement de *E. coli*.

La construction que nous montre la Figure 23.11 permet de suivre la réaction d'épissage dans *E. coli*. L'intron est localisé en un site qui interrompt le dixième

Figure 23.11
En plaçant l'intron de *Tetrahymena* à l'intérieur de la séquence codante de la β -galactosidase, nous pouvons déterminer, par un test simple, sa capacité d'épissage dans *E. coli*. La synthèse de la β -galactosidase est contrôlée en ajoutant un composé qui devient bleu si l'enzyme est présente. La séquence entière est portée par un bactériophage, la présence de plaques bleues indique donc le succès de l'épissage.



codon de la séquence codante de la β -galactosidase. La protéine ne peut donc être traduite efficacement qu'à partir d'un ARN dont l'intron a été épissé.

La survenue de cette réaction dans ce système nous indique que l'épissage peut avoir lieu dans des conditions relativement éloignées de celles prévalant dans *Tetrahymena* ou même *in vitro* ; par exemple, il est probable que les ribosomes se fixent sur l'ARN dès qu'il est synthétisé, avant même qu'il n'ait subi l'épissage. Une interprétation de ce résultat serait que l'auto-épissage peut se produire dans la cellule bactérienne. Une autre possibilité serait que les enzymes bactériennes puissent en fait aider la réaction.

Compte tenu de cet essai, nous pouvons introduire des mutations dans l'intron et voir si elles inhibent la réaction. Les mutations dans les séquences consensus de la classe I, qui rompent leur appariement, arrêtent l'épissage. Cela nous prouve que ces séquences consensus sont importantes. Pour démontrer directement qu'elles fonctionnent par appariement, il serait possible de prouver que les mutations peuvent être annulées par des changements compensatoires restaurant l'appariement initial.

Les mutations ayant lieu dans les séquences consensus correspondantes des introns mitochondriaux de classe I ont des effets similaires. Une mutation dans une séquence consensus peut être réversée par une mutation dans sa séquence consensus partenaire qui restaure ainsi l'appariement ; par exemple, les mutations dans le consensus R peuvent être compensées par des mutations dans le consensus S. Cela suggère que la réaction d'épissage de classe I dépend de la formation d'une structure secondaire entre les paires de séquences consensus à l'intérieur même de l'intron.

UN INTRON POURRAIT CODER POUR UNE PROTÉINE RÉGULATRICE

Dans les mitochondries fongiques, certains introns ont une organisation inhabituelle : *chacun de ces introns possède une séquence codante étendue, dont la traduction peut être indispensable à l'épissage de l'intron dans lequel elle se trouve*. On trouve des introns possédant de telles séquences codantes à la fois dans la classe I et dans la classe II.

Le gène codant pour le cytochrome-b est connu sous les deux noms de *box* et de *cob* et existe dans les deux versions décrites à la Figure 23.12. Les mutations inhibant la synthèse protéique sont localisées dans une série de groupes, qui entrent dans trois catégories.

Des mutations dans les exons affectent directement la protéine cytochrome-b. Tous ces mutants synthétisent un ARNm normal. Les effets des mutations se font sentir au niveau de la traduction, en provoquant des lectures non-sens ou aberrantes du message. Aucune de ces mutations n'en complémente une autre, que ce soit dans le même groupe ou dans un autre. Selon ce critère, elles se situent toutes dans le même gène. Leurs propriétés correspondent exactement à celles que l'on attendrait pour un gène morcelé.

Dans les introns, des mutations agissant en *cis* déterminent les sites nécessaires à l'épissage. Les groupes *box 9* et *box 2* sont situés dans les séquences consensus de l'intron 14 de la classe I ; ils correspondent au R et au S de l'intron de *Tetrahymena*. Ces mutations ne peuvent pas compléter les mutations des autres groupes. Selon ce critère génétique, on ne peut donc pas les distinguer des mutations exoniques. Mais leurs propriétés biochimiques sont différentes : la synthèse de l'ARNm normal est impossible puisque le 14 ne peut être épissé.

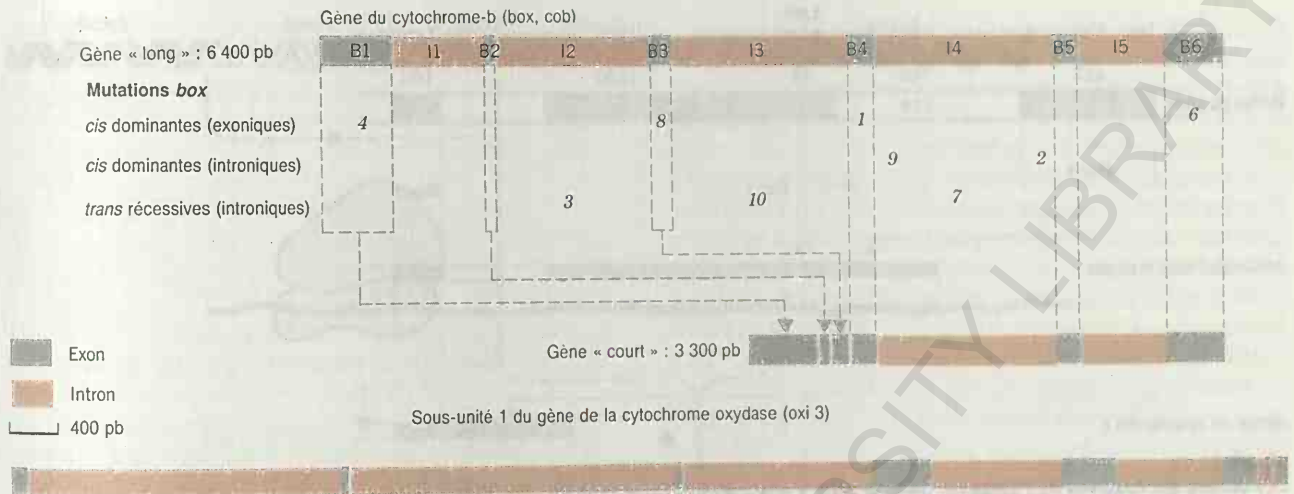
L'existence de cette classe de mutations révèle deux importantes généralités :

- Les mutations dans des sites spécifiques rendent impossible la reconnaissance des jonctions spécifiques d'épissage ; et ces sites peuvent être relativement distants des jonctions réelles elles-mêmes.
- De telles mutations ne peuvent pas se distinguer génétiquement des mutations de la protéine elle-même. (La même impossibilité de discrimination s'applique aux mutations des promoteurs ou des opérateurs et leurs gènes structuraux, mutations du type classique agissant en *cis* ; voir le Chapitre 10.)

Les mutations agissant en *trans* dans les introns caractérisent un nouveau type de fonction impliquée dans l'épissage. Les mutations à l'intérieur de chacun des groupes *box 3*, *box 10* et *box 7*, se comportent comme un groupe de complémentation. Mais, bien qu'elles ne puissent pas compléter d'autres mutations dans le même groupe, elles peuvent compléter des mutations dans tout autre groupe, qu'elles soient de ce type ou de type exonique.

Formellement, cela signifie que chacun de ces trois groupes code pour un produit agissant en *trans*, diffusible, distinct de la protéine cytochrome-b, mais néanmoins nécessaire à sa synthèse. En termes pratiques, chacun de ces trois introns doit contenir des séquences dont la fonction est de coder pour un produit donné, qui a une existence indépendante et joue un certain rôle régulateur dans la production du cytochrome-b.

Chacune de ces mutations entrave la maturation de l'ARNm du cytochrome-b, en empêchant l'épissage de

**Figure 23.12**

Le gène codant pour le cytochrome-b mitochondrial de levure possède une structure mosaïque dans laquelle on trouve trois types de mutations.

Certaines souches de levure possèdent un gène « long », dans lequel la région codante de 1 155 pb, organisée en 6 exons numérotés de B1 à B6, s'étend jusqu'à environ 6 400 pb. Les introns sont identifiés selon leur numérotation dans le gène long, de I1 à I15.

D'autres souches possèdent un gène « court », dont la longueur est approximativement de moitié, dans lequel les quatre premiers exons du gène « long » forment ensemble un exon unique, continu. Les deux formes du gène sont aussi bien exprimées l'une que l'autre.

Les versions « courtes » et « longues » du gène *box* codent pour la même protéine de cytochrome-b de 385 acides aminés. Les mutations sont regroupées, chaque groupe est indiqué par « box » suivi d'un chiffre qui reflète son ordre d'isolement (et non la position sur la carte). Leurs localisations sur la carte indiquent les exons ou les introns dans lesquels ils se situent.

l'intron qui le porte. Un modèle du rôle des fonctions agissant en *trans* nous est proposé par la séquence du gène. Les traits caractéristiques sont présentés sous la forme d'un diagramme, pour l'exemple du I2, dans la **Figure 23.13**. Le premier exon (B1) porte les 139 codons du N-terminal (de 417 pb) du cytochrome-b. Un intron de 765 pb (I1), dont tous les cadres de lecture sont bloqués, sépare le premier exon du second, un très court exon (le B2, de 5 codons). Cet exon est suivi par le second long intron (I2). La caractéristique principale de cet intron est que les premières 840 pb représentent un cadre ouvert de lecture exactement en phase avec celui du cadre de lecture de l'exon précédent.

Les mutations du *box 3* sont toutes situées dans cette région ; elles créent des codons non-sens dans le cadre de lecture ouvert. Cette partie de l'intron peut donc avoir une fonction codante ; et la protéine peut fournir la fonction diffusible impliquée dans l'épissage. Elle est appelée **ARN maturase**. (Le restant de l'intron est bloqué dans tous les cadres de lecture.)

Quelle est la structure de l'ARN maturase ? Une preuve

indirecte nous fait supposer que sa traduction n'est pas initiée à l'intérieur de l'intron, mais a lieu par l'intermédiaire de la lecture du second exon. La matrice adéquate pour cette synthèse serait engendrée par la seule suppression du premier intron, mettant ainsi en contact le premier et le second exon, et créant un cadre de lecture se poursuivant au long du second intron. Le cadre codant est long de 425 codons ; et sa traduction produirait une protéine comprenant les 144 acides aminés du N-terminal du cytochrome-b et les 279 acides aminés codés par l'intron.

Si l'ARN maturase codée par cette séquence est spécifiquement employée à l'excision du second intron, l'action de l'enzyme crée une boucle exerçant un contrôle négatif extrêmement sensible, illustrée à la **Figure 23.14**. L'excision du second intron, en réunissant les deux premiers exons au troisième, interrompt la séquence codante pour la maturase. Ainsi l'activité responsable de l'épissage de l'ARN maturase mène à la perte de la fonction de synthèse de l'enzyme ! Un équilibre sera donc réalisé entre les niveaux des deux formes de l'ARN et de la protéine de l'ARN maturase.

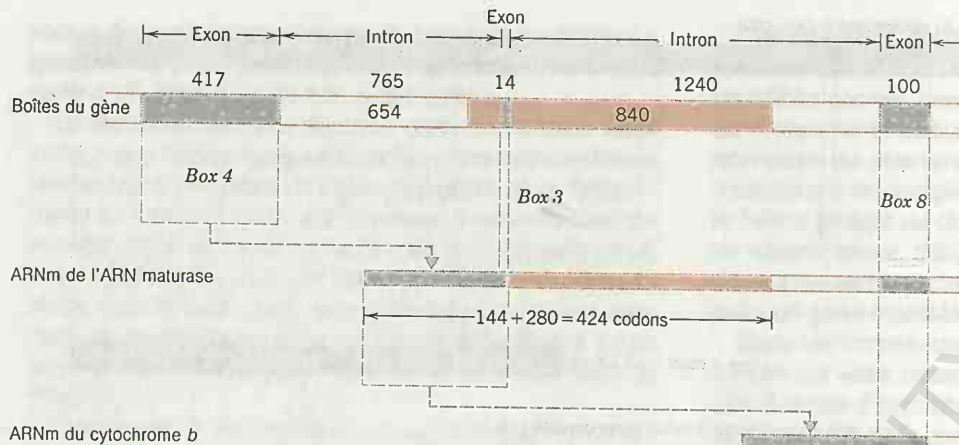


Figure 23.13

Les étapes successives de l'épissage dans le gène « box » pourraient engendrer des ARNm codant pour des protéines chevauchantes. La seule suppression du premier intron engendre l'ARNm de l'ARN maturase ; la suppression du second intron donne ensuite le début de la séquence codante du cytochrome-b.

Dans un autre cas, existent des indices pour qu'une fonction codée par un intron puisse être impliquée dans deux processus d'épissage : les mutants du *box 7* (dans le quatrième intron du gène *box*) sont dénués de la fonction d'épissage, non seulement de l'ARNm du *box*, mais également de l'ARNm de l'*oxi 3*. Les mutations du *box 7* créent des codons non-sens dans le cadre ouvert de lecture de l'intron ; ils peuvent donc bloquer la synthèse d'une autre maturase, celle-là même nécessaire à l'excision de l'intron 14 du gène *box* et d'un intron de l'*oxi 3*.

La présence de séquences de maturase dans les introns des classes I et II pose un paradoxe intéressant au sujet des mécanismes d'épissage. Certains introns de la classe I et de la classe II sont intraductibles ; donc, quelle que soit la fonction remplie par la maturase, elle n'est pas dans ces cas, indispensable. Par ailleurs, les deux types d'intron sont auto-épissables *in vitro* ; la séquence d'ARN possède donc la capacité intrinsèque d'engendrer la structure secondaire/tertiaire adéquate, et de catalyser la rupture et la réunion des liaisons.

Des protéines accessoires sont peut-être nécessaires pour stabiliser la structure. Il est certainement vrai que tout acide nucléique aussi long que ces introns aura des structures alternatives disponibles. Des protéines accessoires pourraient stabiliser la structure nécessaire à l'épissage. De telles fonctions peuvent être fournies totalement ou en partie par la maturase dans le cas d'un intron véhiculant une séquence codante, et par des protéines agissant *in trans* de façon conventionnelle dans le cas des introns dénués de maturase.

L'idée que chaque cadre de lecture d'un intron code pour une maturase spécifique de cet intron (ou tout au moins spécifique d'un petit nombre d'introns), nous fait supposer que son rôle est plus probablement centré sur des phénomènes spécifiques de reconnaissance que sur des activités catalytiques *per se*. Par exemple, la maturase pourrait constituer une « sous-unité marqueur de spécificité » d'un complexe dans lequel un quelque autre composant serait responsable de la rupture et de la création de la présente liaison. L'expérience clé de ce modèle sera l'isolement de la maturase correspondant au cadre ouvert de lecture d'un intron, et la caractérisation de son activité *in vitro*, définissant ainsi l'emplacement de son site de liaison à l'intron, son activité catalytique, si elle en possède une, et dans ce cas laquelle.

Des polymorphismes liés à la présence ou à l'absence des introns sont relativement courants dans la mitochondrie fongique. Ceci est cohérent avec l'idée selon laquelle les introns se seraient formés par insertion dans le gène. L'analyse des recombinaisons croisées mettant en jeu le grand gène d'ARNr de la mitochondrie de levure permet d'éclairer le processus impliqué.

Ce gène possède un intron de classe I qui ne semble pas être nécessaire à la réaction d'épissage, mais qui est lié à une des séquences de la maturase du gène *box*. Cet intron est présent dans certaines souches de levure (appelées ω^+) et absent dans d'autres (ω^-). Les croisements génétiques entre ω^+ et ω^- respectent une polarité : les descendants sont habituellement ω^+ .

Les mutants ne possédant pas cette polarité présentent

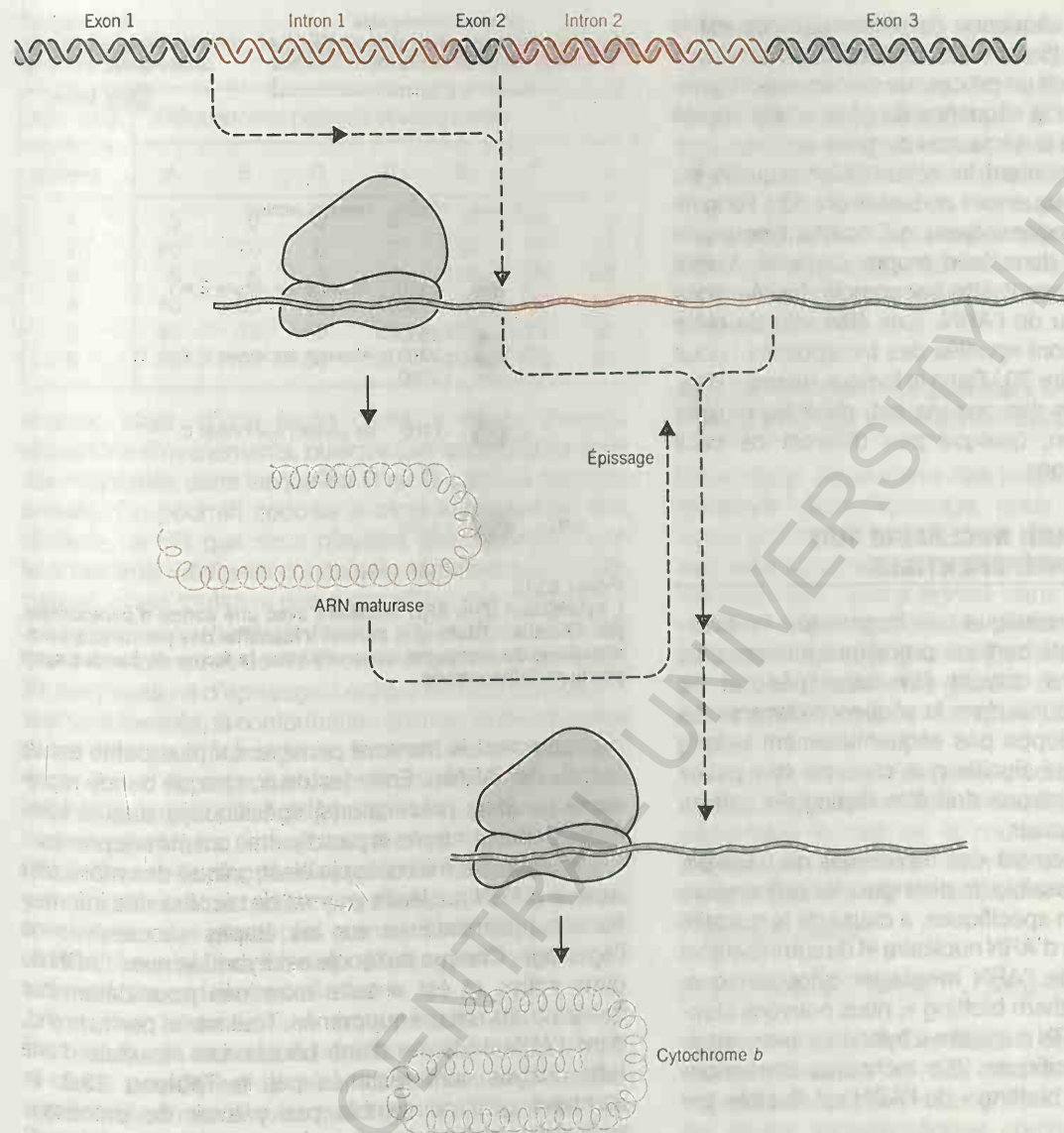


Figure 23.14

Si l'ARN maturase excise spécifiquement l'intron qui code pour elle, une boucle réglera par rétro-contrôle les quantités de la protéine et des deux ARNm.

Les régions ombrées sur l'ADN et l'ARN indiquent des cadres de lecture ouverts, les régions en blanc sont bloquées dans tous les cadres de lecture.

soit des modifications de bases dans une souche ω^- près du site où serait inséré l'intron, soit des mutations dans l'intron bloquant sa traduction. Si nous considérons la souche ω^+ comme donneur, et la souche ω^- comme receveur, apparaît l'idée que la protéine codée par l'intron

dans une souche ω^+ reconnaît le site où il devrait être inséré dans une souche ω^- et de ce fait le rend préférentiellement transmissible.

La protéine codée par l'intron ω est une endonucléase pour laquelle le gène ω^- constitue la cible de la coupure

du double-brin. La séquence de reconnaissance est le site d'insertion de l'intron. La rupture du double-brin amorce probablement un processus de conversion génique au cours duquel la séquence du gène ω^+ est copiée pour se substituer à la séquence du gène ω^- .

Ces résultats renforcent la notion selon laquelle les introns portant les séquences codantes ont dû à l'origine être des éléments indépendants qui codaient pour une fonction impliquée dans leur propre capacité à être excisés de l'ARN, ou peut-être à accomplir des réactions analogues au niveau de l'ADN. (Les éléments de cette dernière catégorie sont appelés des transposons ; nous en parlons au Chapitre 30.) Dans la logique de cette idée, le mode d'utilisation des codons est, dans les régions codantes de l'intron, quelque peu différent de celui existant dans les exons.

L'ÉPISSAGE DE L'ARN NUCLÉAIRE SUIT DES PROCESSUS PRÉFÉRENTIELS

Dans les gènes en mosaïque très fragmentés, l'expression du gène présente certains problèmes intéressants. De nombreux introns doivent être supprimés ; et les exons doivent être réunis dans la séquence correcte. La réaction ne se développe pas séquentiellement le long du précurseur, ce qui signifie que chacune des paires des extrémités des introns doit être distinguée tout au long de l'ARN précurseur.

Avant le développement des techniques de transfert, il était tout à fait impossible de distinguer les précurseurs nucléaires des ARNm spécifiques, à cause de la quantité excessivement petite d'ARN nucléaire et de son inévitable contamination par de l'ARN messager cytoplasmique. Mais, grâce au « Northern blotting », nous pouvons identifier des bandes d'ARN nucléaire s'hybridant avec certaines séquences spécifiques. (La technique correspondante au « Southern blotting » de l'ADN est illustrée par la Figure 16.7.)

Lorsque l'ARN de l'oviducte de poulet est analysé en utilisant une sonde dirigée contre l'ARNm de l'ovomucoïde ou de l'ovalbumine, on obtient dans chaque cas une série de bandes hétérogènes. Cela en soi nous permet de supposer que l'épissage doit se produire selon des étapes définies. (Si les sept introns de chaque gène étaient excisés selon un ordre entièrement aléatoire, il devrait exister plus de 300 précurseurs avec des combinaisons différentes d'introns, et nous ne verrions pas de bandes discontinues.)

La Figure 23.15 nous montre une analyse par « Northern blot » des ARNm précurseurs de l'ovomucoïde. La bande la plus large correspond à la taille du gène, et est

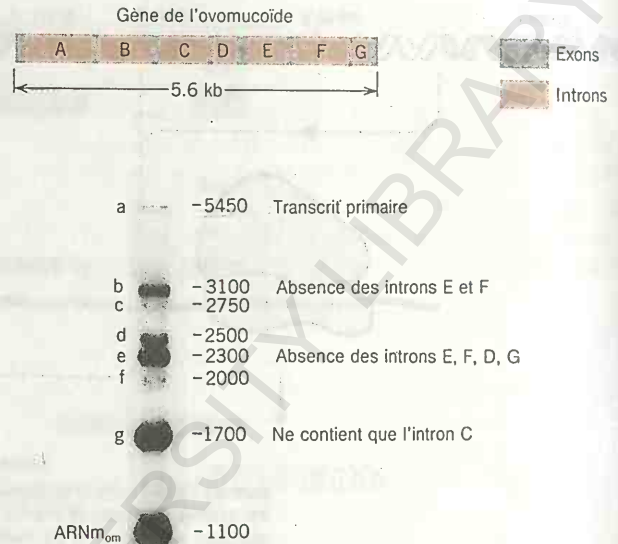


Figure 23.15

L'hybridation d'un ARN nucléaire avec une sonde d'ovomucoïde, par « Northern Blotting », permet d'identifier des précurseurs individualisés du messager observés sous la forme de bandes ayant perdu certains introns.

probablement le transcrit primaire. La plus petite est de la taille de l'ARNm. Entre les deux, chaque bande représente un (des) précurseur(s) spécifique(s) duquel (desquels) certains introns et pas d'autres ont été supprimés.

L'analyse par microscopie électronique des molécules isolées d'ARN nucléaire permet de récolter des informations supplémentaires sur les étapes successives de l'épissage. Chaque molécule est hybridée avec l'ADN du gène entier, et est ensuite examinée pour déterminer quels introns ont été supprimés. Tout intron perdu prend, dans l'ADN, la forme d'une boucle. Les résultats d'une telle analyse sont résumés par le Tableau 23.2. Ils montrent qu'il ne semble pas y avoir de processus *obligatoire*, puisque nous pouvons trouver des intermédiaires dans lesquels différentes combinaisons d'introns ont été supprimées. Cependant, la preuve est faite d'un ou de plusieurs processus *préférentiel(s)*.

Lorsqu'un seul intron a été perdu, c'est virtuellement toujours le E ou le F. Mais l'un ou l'autre peut être perdu le premier. Lorsque deux introns ont été perdus, ce sont encore E et F le plus souvent, mais il existe d'autres combinaisons. L'intron C n'est jamais, ou très rarement, perdu durant l'une des trois premières étapes de l'épissage. D'après ce profil, nous voyons qu'il existe, pour ce processus, un ordre préférentiel au cours duquel E et F sont tout d'abord perdus, puis G et D, B tend à être supprimé ensuite, et A et C sont les derniers à être

Tableau 23.2

Les introns sont excisés de l'ARN nucléaire de l'ovomucoïde selon un ordre préférentiel.

Nbre total d'introns excisés	Fréquence de perte de chaque intron						
	A	B	C	D	E	F	G
1	5	0	0	0	30	60	5
2	20	20	0	2	60	60	25
3	5	5	5	30	100	95	60
4	10	25	35	95	90	90	55
5	40	75	65	85	100	75	60
6	55	100	80	90	100	100	80

excisés. Mais, d'une façon claire, il existe d'autres séquences d'événements, puisque (par exemple), il existe des molécules dans lesquelles D ou G sont les derniers excisés. On pourrait opposer à cette interprétation des résultats, le fait que nous n'ayons pas de preuve que tous ces intermédiaires conduisent réellement à l'ARNm mature ; c'est toutefois une supposition raisonnable.

La conclusion générale induite par cette analyse est que la conformation de l'ARN peut influencer sur l'accessibilité des jonctions d'épissage. Lorsque des introns particuliers sont excisés, la conformation change, et de nouvelles paires de sites d'épissage deviennent disponibles. Mais, étant donné la capacité du précurseur à éliminer ses introns dans des ordres différents, il est possible que des conformations alternatives soient disponibles à chaque étape. Bien entendu, plus la molécule est longue, plus il y a d'options structurales possibles ; et lorsque nous considérons des gènes plus grands, il est difficile de voir comment sont prises les dispositions pour imposer une structure secondaire spécifique.

Tous les intermédiaires sont polyadénylés, ce qui implique que la transcription habituelle est menée à bien, et que le poly (A) est ajouté à, au moins, une proportion appréciable des molécules avant que l'épissage n'ait lieu. Dans ce cas, l'épissage succède à la transcription, et n'est pas concomitant. Cependant, cette succession d'événements n'est pas obligatoire, car l'épissage peut se poursuivre même si la polyadénylation est inhibée (voir Chapitre 24).

Qu'arrive-t-il aux fragments introniques excisés ? Des expériences d'hybridation sur une population d'ARN nucléaires montrent que la concentration en séquences d'introns de l'ovalbumine est à peu près dix fois moins élevée que la concentration en séquences d'exons. Cela signifie probablement que les fragments d'introns libérés sont relativement rapidement détruits. Cela renforce également l'argument selon lequel les précurseurs identifiés représentent l'état d'équilibre de la population nucléaire.

Tous les intermédiaires ayant subi une maturation plus rapide seront présents dans des quantités relativement réduites. Il est donc possible de commettre des erreurs sur les précurseurs identifiés, du fait d'une concentration augmentée en n'importe quel intermédiaire dont la maturation est plus lente. Les résultats de l'expérience de chasse (pulse-chase), où l'on incorpore rapidement un marqueur radioactif à un précurseur primaire que l'on suivra ensuite au cours des diverses étapes de la maturation, nous permettront d'élucider le détail du processus d'épissage.

LES JONCTIONS NUCLÉAIRES D'ÉPISSAGE PEUVENT ÊTRE INTERCHANGEABLES

Pour affiner notre vision des phénomènes moléculaires impliqués dans l'épissage, nous devons prendre en considération la nature des **jonctions d'épissage**, séquences qui jouxtent les sites de coupure et de réunion. Comme nous l'avons signalé dans le Chapitre 18, leurs noms proviennent de leurs positions dans l'intron. La jonction d'épissage gauche (parfois appelée le donneur) est située à l'extrémité gauche de l'intron ; la jonction droite (parfois appelée l'accepteur) est située à l'extrémité droite.

Tous les introns des gènes nucléaires structuraux qui codent pour les protéines obéissent à la loi GT...AG, ainsi dénommée du fait de la courte séquence consensus existant aux jonctions exon-intron. Tel qu'il est décrit dans le Chapitre 18, le consensus est :

Jonction gauche	Jonction droite
↓	↓
Exon... A G G T A A G T ... Intron... Py ₁₀ C A G... Exon	

Les flèches indiquent les liaisons dont l'épissage convertirait le précurseur en ARN mature.

Ces courtes séquences consensus sont apparemment les seules caractéristiques communes aux jonctions d'épissage. Elles signent les frontières exon-intron d'un champ extrêmement étendu d'espèces, allant de l'animal à la levure.

Les courtes séquences consensus ont été reconnues comme des sites impliqués dans l'épissage *in vivo* grâce à l'existence dans ces séquences de mutations ponctuelles qui préviennent la réaction d'excision-épissage. Avec les systèmes *in vitro*, nous pouvons créer des mutations dans ces sites pour montrer qu'ils sont nécessaires à l'épissage. Tout en permettant d'impliquer les séquences consensus dans les phénomènes naturels d'épissage, ces expériences indiquent également que la création de ces sites dans un autre contexte peut fournir une nouvelle

cible au mécanisme d'épissage, alors utilisée comme substitut à la jonction normale.

La reconnaissance de la paire correcte des jonctions d'épissage droite et gauche dépend-elle de chaque séquence en elle-même, ou bien est-elle imposée par une certaine caractéristique, indépendante de l'interaction avec le mécanisme d'épissage ? En d'autres termes, chaque paire correcte est-elle marquée d'une certaine façon par une caractéristique unique de séquence ou de structure, ou bien toutes les jonctions gauches sont-elles équivalentes d'un point de vue fonctionnel, et toutes les jonctions droites indifférenciables de la même manière ?

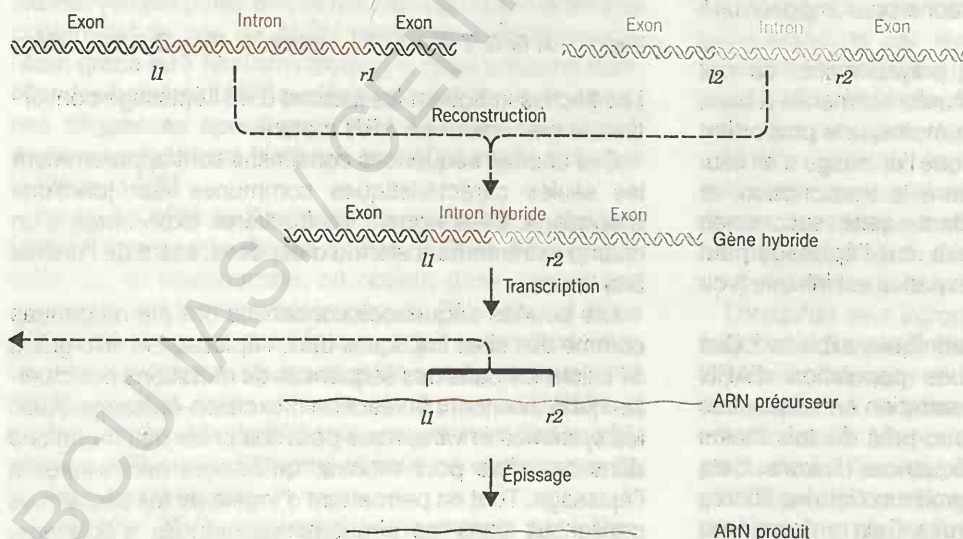
D'après les séquences situées aux jonctions d'épissage (et autour), nous savons qu'il n'existe pas de complémentarité entre les jonctions correspondantes gauches et droites, ce qui exclut l'idée qu'elles puissent être directement réunies par appariement. Cela laisse la possibilité que les paires individuelles de séquences de jonctions, puissent être spécifiquement reconnues par des protéines ou par des séquences d'ARN. Des schémas de ce type peuvent s'appliquer aux mitochondries de levure ou aux gènes nucléaires d'ARNr, où seul un petit nombre de paires de jonctions est impliqué, mais semble plus difficile à envisager dans le noyau, le nombre des paires de jonctions y étant beaucoup plus important.

Par voie de conséquence, on peut se demander s'il existe une spécificité tissulaire de l'épissage de l'ARN. Le mécanisme d'épissage est-il le même pour tous les gènes, quelle que soit la circonstance, de telle manière

que la reconnaissance des jonctions appropriées ne dépende que de l'ARN et du mécanisme commun d'épissage ? Ou bien, la capacité d'épisser un ARN nucléaire spécifique n'est-elle présente que dans les cellules à l'intérieur desquelles il est exprimé ? Ces questions peuvent trouver une réponse grâce à l'étude des authentiques ARN précurseurs dans des situations cellulaires nouvelles, et à la construction de gènes synthétiques possédant des combinaisons de jonctions d'épissage nouvelles.

Pour insérer des gènes dans des virus eucaryotes, il est possible d'utiliser les mêmes techniques que celles décrites au Chapitre 16 pour le clonage de séquences étrangères dans les phages bactériens ou dans les plasmides. En réalisant, grâce à des enzymes de restriction, des coupures appropriées, et en réunissant ensuite les fragments, la globine, l'insuline et d'autres gènes de sources variées ont été insérés dans des vecteurs viraux eucaryotes. Un des vecteurs les plus courants est le virus SV40 du singe qui peut se maintenir dans les cellules de différents mammifères. Les nouveaux gènes peuvent être insérés dans l'unité de transcription, soit précoce, soit tardive, du virus. Leur expression passe par la formation d'un ARN précurseur qui, habituellement, contient des séquences virales en son extrémité 5', et se poursuit par le gène étranger.

Cela nous apporte deux informations importantes. En premier lieu, un épissage conforme ne dépend pas de l'intégrité de l'ARN précurseur naturel, puisqu'un gène



l et *r* indiquent les jonctions d'épissage gauches et droites

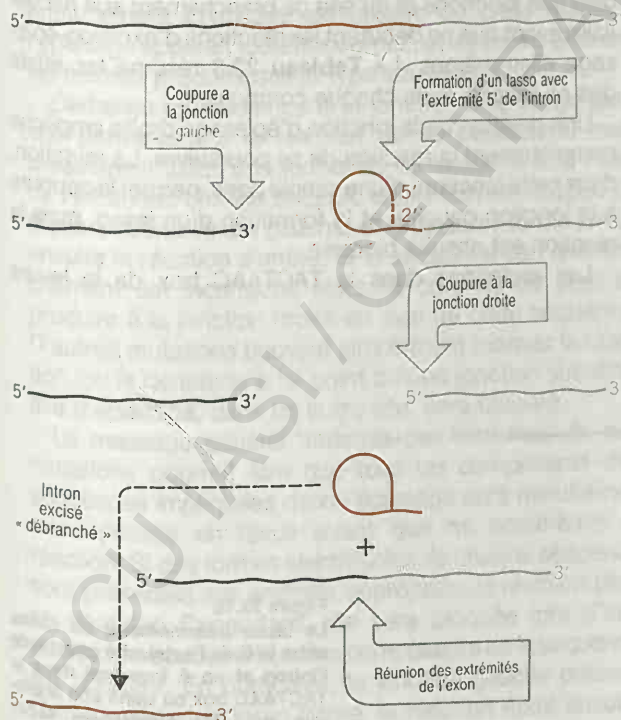
Figure 23.16

L'épissage se produit normalement dans le transcrit d'une unité construite en unissant les exons de deux gènes différents.

étranger (ou parfois seulement une partie de ce gène) peut être l'objet d'une maturation correcte dans le contexte de séquences virales. En second lieu, l'information véhiculée par la séquence est conforme au consensus favorisant l'épissage, à partir du site habituel d'expression, dans des cellules provenant d'un tissu, et/ou d'une espèce différente. Les signaux impliqués dans l'épissage doivent donc être extrêmement bien conservés.

Nous avons poussé cette approche, par reconstruction, un peu plus avant, en fabriquant des gènes synthétiques dans lesquels un exon provenant d'un gène authentique est lié à un exon provenant d'un gène différent. Le diagramme de la **Figure 23.16** illustre cela. Au cours d'une expérience, le premier exon de l'unité précoce de transcription du SV40 a été associé au troisième exon de la β -globine de souris. L'intron hybride a été très correctement excisé. Ainsi la jonction gauche d'un intron SV40 (11) peut être épissée à la jonction droite d'un intron (r2) de β -globine de souris.

Figure 23.17
L'épissage *in vitro* implique une coupure à la jonction gauche, suivie de la formation d'un lasso dans lequel l'extrémité gauche de l'intron est unie, par une liaison 5'-2', à un site proche de l'extrémité droite de l'intron. Les étapes suivantes impliquent une coupure à la jonction droite et une liaison covalente entre les exons. L'intron excisé est ensuite « débranché ».



Cela soulève un paradoxe. Toutes les jonctions d'épissage gauches sont probablement regardées comme équivalentes par les systèmes d'épissage, de même que les jonctions droites. *En principe, toute jonction d'épissage gauche doit être capable de réagir avec toute jonction d'épissage droite.* Mais, dans les circonstances les plus courantes, l'épissage ne se produit qu'entre la jonction gauche et la jonction droite du même intron. *Comment la reconnaissance des jonctions d'épissage est-elle restreinte, de sorte que seules les jonctions gauches et droites du même intron soient épissées ?*

Les modèles mettant en cause la conformation de l'ARN deviennent moins plausibles, si l'on considère l'épissage apparemment normal des gènes hybrides. Une des possibilités serait que le système d'épissage agisse par étapes. Ayant reconnu une jonction, l'enzyme est tenue de passer en revue l'ARN dans le sens adéquat jusqu'à ce qu'elle rencontre une jonction de l'autre nature. Cela limiterait l'épissage à des jonctions adjacentes.

Ce modèle pourrait expliquer les cas où les jonctions sont définies de manière unique et interagissent (bien que certains introns soient en fait très longs). Cela n'explique pas l'existence de modes alternatifs d'épissage, comme ceux présents dans le SV40, où une jonction gauche classique peut être épissée avec plus d'une jonction droite. (Bien entendu, dans tout cela nous supposons que l'épissage, habituellement, s'effectue correctement ; mais nous ne possédons aucune donnée nous indiquant selon quelle fréquence des épissages erronés se produisent en réalité.)

UN « SPLICEOSOME » NUCLÉAIRE ENGENDRE UN LASSO

Pour éclaircir définitivement le mécanisme de l'épissage, il faut des systèmes *in vitro* où les introns pourraient être excisés des ARN précurseurs. Grâce à de tels systèmes, il est possible de caractériser les intermédiaires et de tester l'influence de composants donnés sur la réaction.

Des extraits nucléaires de cellule HeLa peuvent épisser des ARN précurseurs purifiés, ce qui nous démontre que l'acte d'épissage n'est pas lié au processus de transcription. De même, l'épissage n'est pas lié aux modifications de l'ARN. Les ARN de globine qui n'ont pas de queues poly (A) sont parfaitement bien épissés ; bien que les ARN pourvus d'une coiffe (cap) soient plus efficacement épissés que les ARN n'en possédant pas, il n'y a pas d'autre différence dans leur réaction.

L'épissage *in vitro* se produit en deux étapes, qu'illustre le schéma de la **Figure 23.17**. A la différence de la

réaction d'auto-épissage qui ne requiert pas d'énergie, l'épissage nucléaire nécessite de l'ATP.

Lors de la première étape, une coupure est réalisée à l'extrémité gauche de l'intron, dégageant, sous la forme de molécules séparées d'ARN, l'exon gauche et la molécule intron-exon droit. Les deux espèces d'ARN sont probablement maintenues ensemble par des protéines impliquées dans l'épissage.

L'exon gauche est libéré sous la forme d'une molécule linéaire, mais la molécule intron-exon droit n'est pas linéaire. La terminaison 5' engendrée à l'extrémité gauche de l'intron se lie, par une liaison 5'-2', au A de la séquence CTGAC, située à environ 30 bases en amont de l'extrémité droite de l'intron. La liaison produit un « lasso ».

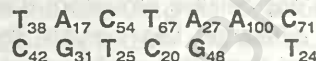
Lors de la seconde étape, une coupure à la jonction d'épissage droite dégage l'intron libre sous forme d'un lasso, tandis que l'exon droit séparé est épissé à l'exon gauche. Le lasso est ensuite « débranché » pour donner un intron linéaire excisé.

Les trois seules séquences nécessaires à l'épissage sont donc de courtes séquences consensus situées aux jonctions d'épissage gauche et droite et à la séquence où l'embranchement se formera. Outre la notion que la majeure partie de la séquence d'un intron peut être déletée sans que cela nuise à l'épissage, cela nous indique qu'une conformation spécifique de l'intron n'est pas nécessaire (et cela contraste avec les introns de type I, auto-épissables).

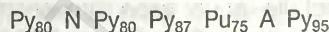
Le site d'embranchement est apparenté à une séquence conservée dans les introns nucléaires de levure ; il est situé en amont, à proximité de la jonction droite d'épissage. Des mutations ou des délétions de cette séquence rendent l'épissage impossible dans la levure ; l'insertion d'une séquence synthétique à l'intérieur d'un intron provoque des perturbations dans l'épissage. La séquence consensus conservée est appelée la **TACTAAC box**. Elle est complémentaire de la jonction d'épis-

sage gauche, mais une analyse par mutations suggère qu'en fait elle ne s'apparie pas avec la jonction.

La séquence cible de l'embranchement chez les eucaryotes supérieurs n'est pas bien conservée, mais semble se conformer dans l'ensemble au consensus suivant :



Bien que d'une manière globale il n'y ait habituellement pas de choix précis d'une base spécifique à chaque position, il y a une préférence évidente pour les purines au détriment des pyrimidines, donnant le consensus :



qui est une forme plus générale de la TACTAAC box. Le lasso se forme au niveau du A, en sixième position de la séquence TACTAAC ; seule cette base doit être entièrement conservée chez les eucaryotes supérieurs. Le point d'embranchement se situe 18 à 40 nucléotides en amont de la jonction d'épissage droite.

La liaison que forme le lasso s'étend de la position 5' du G invariant qui était à l'extrémité gauche de l'intron, à la position 2' du A invariant du site d'embranchement. La structure de l'embranchement est dessinée à la **Figure 23.18**.

Bien que la réaction se fasse en étapes enzymatiques discontinues, il est possible qu'une reconnaissance globale des jonctions et du site de branchement soit nécessaire avant que ne débutent les réactions d'excision-épissage elles-mêmes. Le **Tableau 23.3** résume les effets des mutations dans chaque composant.

Une délétion de la jonction d'épissage droite empêche complètement la réaction de se poursuivre. La mutation, dans cette jonction, d'une simple base, permet la coupure à la jonction gauche, et la formation d'un lasso, mais la réaction est ensuite bloquée.

Les mutations dans la TACTAAC box de la levure

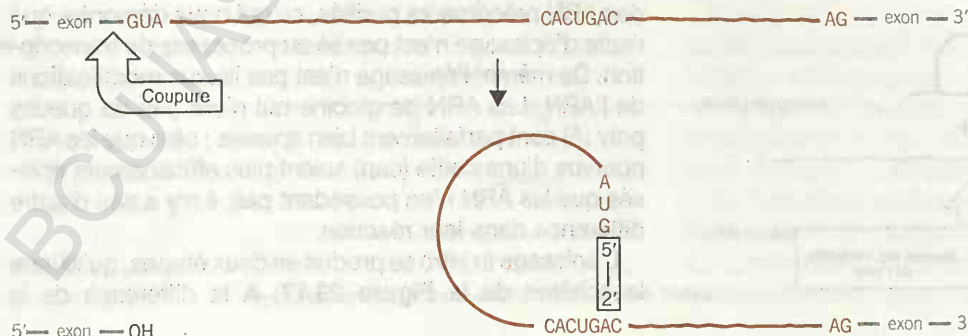


Figure 23.18

Le lasso intermédiaire se forme entre le G de l'extrémité gauche de l'intron et un A invariant dans la TACTAAC box ou dans son analogue chez les eucaryotes supérieurs.

Tableau 23.3

Des mutations dans les séquences impliquées dans l'épissage peuvent bloquer diverses étapes de la réaction.

Système	Consensus	Mutation	Effet
Mammifère »	jonction droite (AG)	délétion GG	pas de réaction bloquée au lasso
Levure Mammifère	TACTAAC box site d'embranchement	TACTACC délétion	pas de réaction utilisation d'un embranchement cryptique
Levure Mammifère	jonction gauche (GT)	GTAPyAT AT	pas de réaction bloquée au lasso

empêchent toute réaction de se produire. Ainsi le site cible de la formation de l'embranchement doit être reconnu pour que soit réalisée la coupure de la jonction gauche de l'intron qui libère l'extrémité 5' de l'intron.

La séquence du point d'embranchement chez les eucaryotes supérieurs fait preuve d'une flexibilité plus importante. Si la séquence habituelle de l'embranchement est inactivée, une séquence apparentée et proche sera, la plupart du temps, utilisée (bien qu'elle soit moins efficace que la séquence authentique). La proximité de la jonction d'épissage droite semble être importante, puisque le site cryptique est toujours situé de 22 à 37 nucléotides de la jonction. Lorsqu'une séquence d'embranchement cryptique est utilisée de cette manière, l'épissage apparaît toutefois normal, et les exons donnent les mêmes produits que le type sauvage.

Certaines mutations dans la jonction d'épissage gauche permettent à la réaction de se poursuivre, mais uniquement jusqu'à la formation du lasso. Si le G initial de l'intron est changé en un A, la jonction est coupée de façon à libérer l'exon gauche, et le lasso se forme ; mais ensuite la réaction s'arrête. Si la séquence de l'embranchement est incorrecte, alors, la coupure ne peut se produire à la jonction droite en aval de cette séquence. D'autres mutations peuvent simplement bloquer la réaction (ou la dénaturer à tel point qu'une jonction substitutive d'épissage, dans un autre site, sera utilisée).

Le message général transmis par les effets de ces mutations pourrait être que tous les composants des séquences impliquées dans l'épissage sont minutieusement passés en revue avant que ne commence la réaction. Si des formes identifiables de chaque séquence sont présentes aux endroits appropriés, la réaction peut se dérouler. Cependant, elle sera bloquée lors d'une étape consécutive si l'un des composants de la séquence est aberrant. En particulier, la jonction gauche pourrait être coupée et le lasso formé, la réaction étant ensuite bloquée si la jonction droite n'est pas fonctionnelle.

L'existence d'un complexe appelé « **spliceosome** » peut expliquer ces résultats. Isolé à partir des systèmes d'épissage *in vitro*, ce complexe comprend une particule de ribonucléoprotéine de 50-60S. Un « complexe de pré-épissage » est assemblé avant qu'aucune réaction n'ait lieu. Sa formation requiert la reconnaissance de la jonction d'épissage gauche et de la séquence d'embranchement ; il se forme moins efficacement si la jonction d'épissage droite est absente. La **Figure 23.19** nous offre un modèle des étapes de l'épissage.

Le « spliceosome » se forme sur un ARN précurseur intact, et passe par un état intermédiaire dans lequel il contient la molécule linéaire de l'exon gauche isolé et l'ensemble lasso-intron-exon droit. Nous ne trouvons, dans le complexe, que peu du produit exon gauche épissé-exon droit, ce qui nous fait supposer que la coupure de la jonction droite et la ligation sont immédiatement suivies de la libération du produit.

Les lassos ne sont pas propres à l'épissage nucléaire, nous les trouvons aussi dans la maturation *in vitro* des introns mitochondriaux de classe I ou II. Les caractéristiques de leur formation sont quelque peu différentes dans les deux cas.

Dans le cas des introns auto-épissables de classe I, les points d'embranchement qui créent les lassos semblent se former en diverses positions à l'intérieur de l'intron. Le résidu guanine nécessaire à l'auto-épissage, qui s'intègre à l'extrémité 5' de l'intron (voir la **Figure 23.7**), semble ne pas être incorporé au lasso. Il est possible que la formation du lasso soit un sous-produit de la réaction auto-catalytique, et ne reflète pas le processus naturel.

Lorsque les introns de classe II sont excisés *in vivo*, ils donnent naissance à des molécules circulaires relativement stables. Ces molécules peuvent, toutefois, être, en fait, des lassos avec de courtes queues. On a pu mettre au point une réaction *in vitro*, au cours de laquelle les introns de classe II peuvent être éliminés par une réaction

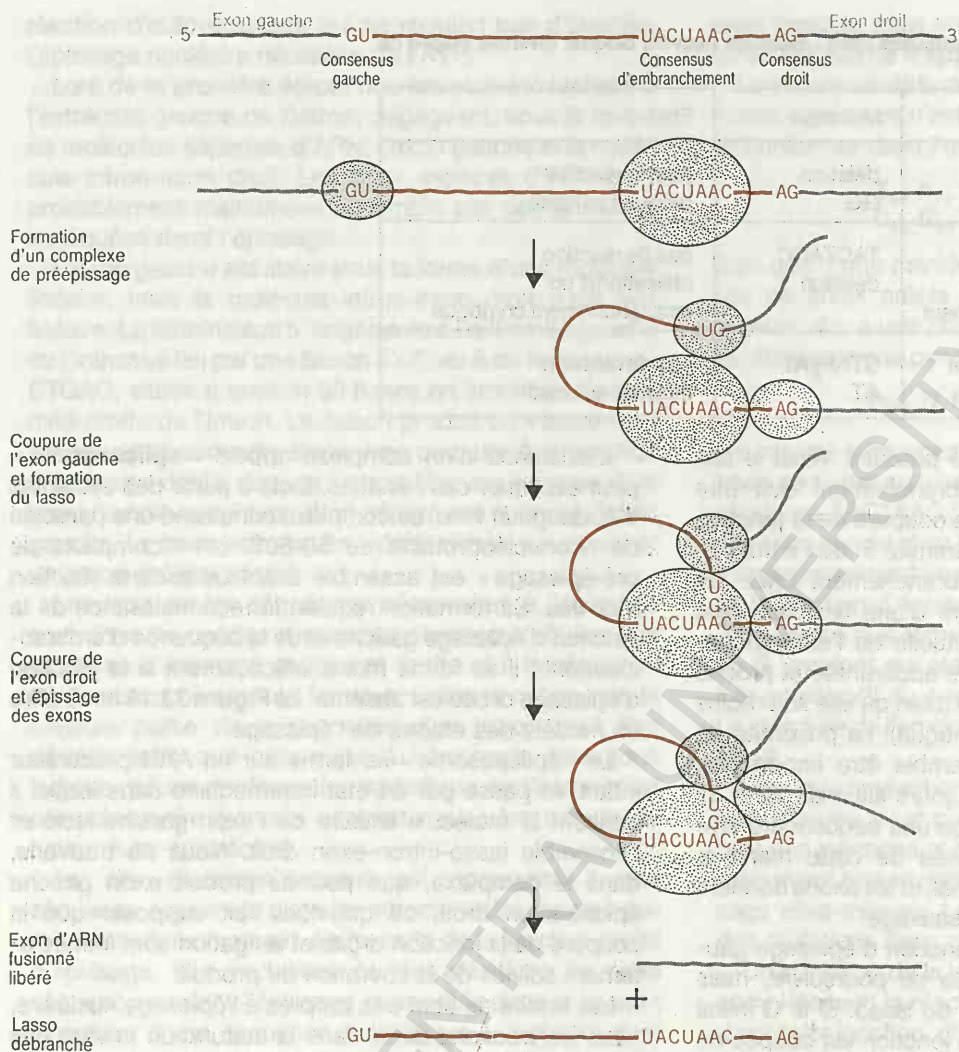


Figure 23.19

La formation d'un spliceosome implique l'interaction des composants qui reconnaissent les séquences consensus ; la réaction d'épissage se déroule par étapes discontinues.

d'auto-épissage. La réaction est autonome : elle ne dépend pas de la présence d'une guanine. Il apparaît que l'excision donne directement naissance au lasso, qui ensuite engendre par coupure une forme linéaire. La forme linéaire ne réagit plus ultérieurement.

LES snARN SONT-ILS IMPLIQUÉS DANS LES RÉACTIONS D'EXCISION-ÉPISSAGE ?

De quelle manière les jonctions d'épissage sont-elles reconnues ? Nous pouvons identifier deux principaux systèmes. Une enzyme pourrait elle-même spécifiquement reconnaître une paire de séquences consensus. Ou bien un ARN pourrait s'apparier avec elles pour créer une

structure secondaire reconnue par une enzyme. En effet, la plupart des enzymes de maturation de l'ARN semblent plutôt reconnaître des conformations que des séquences.

Quelle pourrait être l'origine des séquences complémentaires aux jonctions d'épissage ?

- Ce pourrait être des séquences situées à un autre emplacement dans le même transcrit, impliquant un mécanisme analogue à celui suggéré pour les introns mitochondriaux de levure et les gènes d'ARNr nucléaires (comme le montre la séquence-guide interne à la Figure 23.10).
- Une autre idée serait qu'une (ou des) molécule(s) d'ARN indépendantes puisse(nt) s'apparier au précurseur non épissé.

Il existe un candidat prêt à jouer le rôle de l'ARN agissant en *trans*. Le noyau et le cytoplasme des cellules eucaryotes contiennent de nombreuses espèces de petits ARN. Leur taille varie entre 100 et 300 bases, et leur quantité entre 10^5 et 10^6 molécules par cellule. Certains sont synthétisés par l'ARN polymérase III, d'autres par l'ARN polymérase II, parmi lesquels certains produits sont recouverts d'une coiffe comme les ARNm.

Ceux qui sont circonscrits au noyau sont appelés **petits ARN nucléaires** (snARN = small nuclear RNAs) ; ceux que l'on trouve dans le cytoplasme sont appelés **petits ARN cytoplasmiques** (scARN = small cytoplasmic RNAs). A l'état naturel, ils existent sous forme de particules de ribonucléoprotéines (snRNP et scRNP). Dans le langage courant, on les appelle parfois *snurps* et *scyrps* (snurps : small nuclear ribonucleoproteins ; scyrps : small cytoplasmic ribonucleoproteins). Parmi les divers snARN on trouve certaines analogies de séquence ; les réactions des snRNP avec des anticorps nous font supposer que certaines d'entre elles pourraient avoir des sous-unités protéiques communes.

Plusieurs des snRNP sont impliquées dans l'épissage. Différents snARN possèdent des séquences complémentaires aux jonctions d'épissage gauche et droite et à la séquence de l'embranchement.

Le U1 est un des snARN les plus importants, il est présent dans les cellules animales, chez les oiseaux et les insectes. Le snRNP humain U1 contient 8 protéines et l'ARN U1. La structure secondaire probable du snARN humain U1 est dessinée à la **Figure 23.20**. Les 11 nucléotides de l'extrémité 5' sont simple-brins et comprennent une portion complémentaire de la séquence consensus située à la jonction gauche de l'intron.

La longueur de la zone de complémentarité dans les jonctions réelles est habituellement de 4 à 6 pb. La particule intacte de snRNP U1 peut se lier *in vitro* à une jonction d'épissage gauche. La liaison est une propriété de la particule entière ; l'ARN U1 purifié ne peut se lier.

La preuve de l'implication du snARN U1 dans l'épissage est fournie par les effets des anticorps dirigés contre la particule de ribonucléoprotéine, qui reconnaît probablement une ou plusieurs sous-unités protéiques. L'épissage *in vitro* est inhibé par une incubation en présence d'anticorps anti-snRNP U1, la réaction ne pouvant se dérouler si le système est épuisé en snRNP U1. En fait, la suppression des nucléotides de l'extrémité 5' du snARN U1 affecte l'épissage *in vitro*. La formation du spliceosome dépend de la présence du snRNP U1.

La jonction d'épissage droite est probablement reconnue par la snRNP U5. Nous ne savons pas si, pour fonctionner, la snRNP utilise son composant d'ARN pour

reconnaître, en s'y appariant, la séquence consensus. Il n'y a pas de région de complémentarité disponible dans la forme simple-brin. Les composants protéiques de la « snurp » sont peut-être impliqués.

Le snARN U2 comprend des séquences complémentaires d'un site consensus, qui, dans ce cas, est le site d'embranchement. Par immunoprécipitation avec des anticorps dirigés contre la snRNP U2, il est possible de mettre en évidence un complexe entre les séquences du snARN U2 et celles de l'intron comprenant le site d'embranchement. U1 et U2 peuvent toutes deux être impliquées dans l'étape initiale de l'épissage, puisque leur inactivation empêche la coupure à la jonction gauche et la formation du lasso. D'après la **Figure 23.19**, le snRNP U1 pourrait faire partie du composant se liant au consensus GU, et la snRNP U2 faire partie du composant se liant au consensus UACUAAC. Les deux snARN U semblent être nécessaires à la formation du spliceosome.

Une autre « snurp » contient deux snARN, U4 et U6, qui semblent également être impliqués dans l'épissage. La particule « snurp » est un composant du nucléosome, mais nous ignorons sa fonction.

Nous pouvons penser que les particules de snRNP participent à la construction de la structure du spliceosome. Comme le ribosome, le spliceosome peut dépendre des interactions ARN-ARN, de même que des interactions protéine-ARN et protéine-protéine. Il peut être nécessaire pour certaines réactions impliquant les snRNP que leurs ARN s'apparient directement avec des séquences de l'ARN en train d'être épissé ; d'autres réactions peuvent nécessiter la reconnaissance entre les snRNP ou leurs protéines et d'autres composants du spliceosome.

LES RÉACTIONS D'ÉPISSAGE ONT-ELLES DES POINTS COMMUNS ?

Les caractéristiques des différents types de réactions d'épissage sont résumées par le **Tableau 23.4**. Nous pouvons dénombrer au moins quatre groupes de réactions, que l'on distingue selon les facteurs qu'elles requièrent *in vitro*, et les intermédiaires qu'elles engendrent.

Il ne se dégage aucune caractéristique commune aux deux formes de réactions d'auto-épissage (représentées par les introns de classes I et II), si ce n'est leur autonomie. Dans chaque cas, l'information nécessaire à la réaction est fournie par la séquence intronique (bien que la réaction soit en réalité favorisée, *in vivo*, par les protéines).

L'épissage nucléaire des ARNm précurseurs fournit un contraste marqué. Au cours de cette réaction, seules de

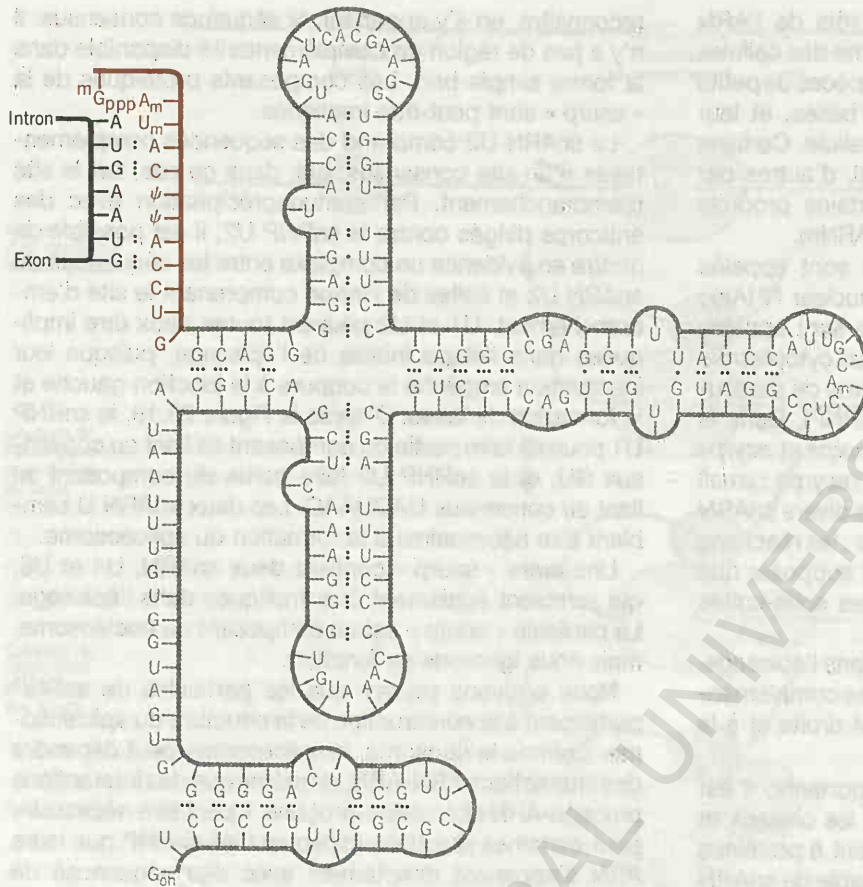


Figure 23.20

Le snARN humain U1 possède une structure secondaire dans laquelle l'extrémité 5' terminale est simple-brin et complémentaire de la séquence consensus d'épissage gauche.

Tableau 23.4

In vitro, les réactions d'épissage possèdent des caractéristiques différentes :

	ARNt nucléaire de levure	ARNr tetrahymena Intron de classe I	Intron de classe II	ARNm précurseur nucléaire
Séquences requises	conformation intron/exon	quatre courts consensus (et guide interne ?)	inconnues	jonctions d'épissage et séquence d'embranchement
Composants exogènes	endonucléase et ligase	un nucléotide à guanine	aucun aucun	spliceosome incluant des snRNP U
Énergie requise	ATP	aucune	aucune	ATP
Formes intermédiaires	moitié d'ARNt	transfert direct d'un phosphoester	lasso	lasso

très courtes séquences consensus sont nécessaires, le reste de l'intron semble être sans importance, et la réaction exige un complexe de spliceosome.

Les introns de classe II ont pourtant en commun avec les introns nucléaires l'utilisation d'un lasso en tant qu'intermédiaire. La formation du lasso permet d'observer un intéressant parallèle avec l'implication de la position 2' dans l'épissage de l'ARNt et la séquestration de l'extrémité 5' de l'ARNr précurseur du *Tetrahymena*. Mais pourquoi l'épissage devrait-il se dérouler par l'intermédiaire d'une structure mettant en jeu l'extrémité d'un intron, qui sera détruite, et un site interne à cet intron ? Peut-être est-il nécessaire de séquestrer l'extrémité 5' libre de l'intron pour ne pas qu'elle bloque la réaction.

Nous pourrions formuler l'hypothèse que des relations entre les mécanismes de ces différents types d'épissage seront découvertes lorsque nous serons à même de définir les fonctions des snARN. Leur rôle en tant que facteurs agissant en *trans* dans l'épissage nucléaire est peut-être analogue au rôle joué par les séquences internes agissant en *cis* des introns mitochondriaux de classe I (et probablement de classe II).

L'implication des snRNP dans l'épissage pourrait n'être qu'un exemple de leur engagement dans les réactions de maturation de l'ARN. Diverses snRNP sont impliquées dans la polyadénylation et dans la production d'extrémités 3' natives dans l'ARNm des histones du *Xenopus* (voir le Chapitre 24). Plus particulièrement, si l'on tient pour acquis le fait que le précurseur de l'ARNr de *Tetrahymena* est auto-épissable, et que l'ARN de la ribonucléase P possède une activité catalytique, on peut raisonnablement considérer que les réactions ARN-ARN peuvent être capitales dans la plupart ou dans tous les phénomènes de maturation de l'ARN.

Ces réactions peuvent prendre des formes variées, mettant en jeu des conformations spécifiques à l'intérieur d'une molécule indépendante (les sites impliqués agissent ainsi en *cis*), ou reposent sur des interactions agissant en *trans* entre le substrat et un autre ARN. L'existence de ces interactions pourrait, en fait, être le vestige le plus important du développement originel des acides nucléiques en tant que molécules héréditaires. Il ne serait pas étonnant que les molécules de snARN se révèlent avoir des fonctions de type catalytique dans l'épissage et dans d'autres réactions de maturation.

Nous avons traité l'épissage en fonction des réactions connues chez les eucaryotes. On a également trouvé des gènes morcelés dans l'archéobactérie, mais nous ne connaissons rien des mécanismes impliqués dans leur expression. Tous les gènes des eubactéries sont continus. Cependant, un gène morcelé a été identifié dans le phage T4, qui s'exprime dans l'eubactérie *E. coli*.

La séquence codante du gène de la thymidylate synthétase (*td*) n'est interrompue que par un seul intron, supprimé par un processus d'épissage. Tel que nous l'avons défini jusqu'à présent, le système du phage semble représenter une réaction d'épissage de classe I. L'intron contient les quatre éléments conservés typiques, avec des exemples parfaits du 9L et du 2, et des exemples relativement frustrés du P et du Q (voir le Tableau 23.1). L'intron contient une phase codante qui est sensée représenter un polypeptide de 245 acides aminés, bien que des données actuelles montrent qu'en fait il pourrait ne pas être exprimé. Une réaction d'auto-épissage se produit *in vitro*, engendrant les deux formes linéaires et circulaires de l'intron excisé. Nous ne savons pas encore si elle se déroule par l'intermédiaire de la réaction de transfert d'un phosphoester comme pour les introns de classe I.

Si la réaction se déroule comme les autres réactions d'auto-épissage *in vivo*, elle aura besoin de l'intervention des protéines agissant en *trans*. Ces produits doivent être codés par le phage ou la bactérie. Il devrait donc être possible d'isoler les mutants ne pouvant pas épisser l'ARN. Grâce aux fonctions perturbées chez ces mutants, il devrait être possible de caractériser la gamme complète des produits nécessaires à l'épissage bactérien ; ainsi nous serions à même de déterminer si cette réaction est une déviation d'autres réactions ou si elle utilise des fonctions particulières. De plus, les fonctions d'épissage ayant subi des mutations, il devrait être possible de déterminer si des gènes autres que celui de la thymidylate synthétase de T4 doivent, pour être exprimés, subir un épissage.

Considéré en fonction de l'évolution et du mécanisme, l'épissage a été interprété comme une réaction *intramoléculaire*, se ramenant essentiellement à une délétion contrôlée des séquences de l'intron au niveau de l'ARN. L'épissage de l'ARN ne peut-il être utilisé comme une réaction *intermoléculaire* afin de relier des exons de différentes molécules d'ARN ? Nous savons que toute réaction de ce type doit être pour le moins inhabituelle, car si elle était répandue les exons d'un gène seraient capables de se compléter génétiquement entre eux, au lieu d'appartenir à un groupe unique de complémentation.

Une expérience visant à étudier l'épissage en *trans* est illustrée à la **Figure 23.21**. Des séquences complémentaires ont été introduites dans les introns de deux ARN différents. L'appariement entre les séquences complémentaires devrait créer une molécule en forme de H. Cette molécule pourrait être épissée en *cis*, pour connecter des exons reliés de façon covalente par un intron, ou bien elle pourrait être épissée en *trans*, pour connecter les exons des molécules juxtaposées d'ARN. Les deux réactions se produisent *in vitro*.

Bien que la présence de séquences complémentaires

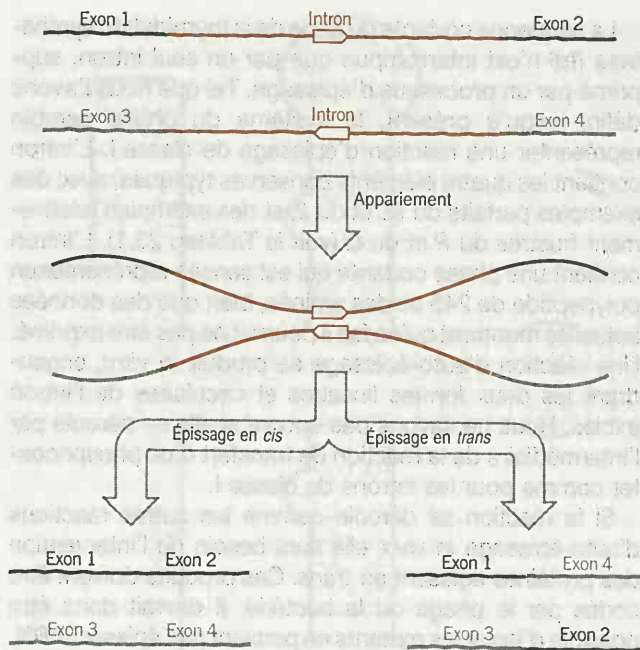


Figure 23.21

L'épissage en *trans* se produit *in vitro* entre deux molécules d'ARN dont les introns possèdent des séquences complémentaires leur permettant de s'apparier.

dans les introns augmente considérablement l'apparition de l'épissage en *trans*, une (certaine) réaction se produit *in vitro* rien qu'en mélangeant deux exons, l'un lié à la moitié gauche d'un intron, l'autre à la moitié droite. Le mécanisme le plus probable est qu'un spliceosome reconnaît séparément les jonctions gauches et droites de l'intron, et s'assemble autour des deux molécules. La survenue de cet épissage en *trans* rend improbable le mécanisme du balayage ; selon ce mécanisme, l'appareil d'épissage reconnaîtrait la jonction gauche, puis rechercherait une jonction droite en balayant une molécule d'ARN intacte.

L'épissage en *trans* ne se produit-il jamais *in vivo* ? Il pourrait expliquer la présence fréquente d'une séquence leader de 35 bases à l'extrémité de nombreux ARNm du trypanosome. Cette séquence leader n'est pas codée en amont des unités de transcription indépendantes. Au lieu de cela, elle est transcrite, à partir d'une unité répétée située dans une autre partie du génome, en un ARN indépendant, possédant en son extrémité 3' des séquences supplémentaires. La Figure 23.22 nous montre que le produit porte la séquence leader de 35 bases, suivie d'une séquence de jonction d'épissage gauche. Les séquences codant pour les ARN messagers portent une

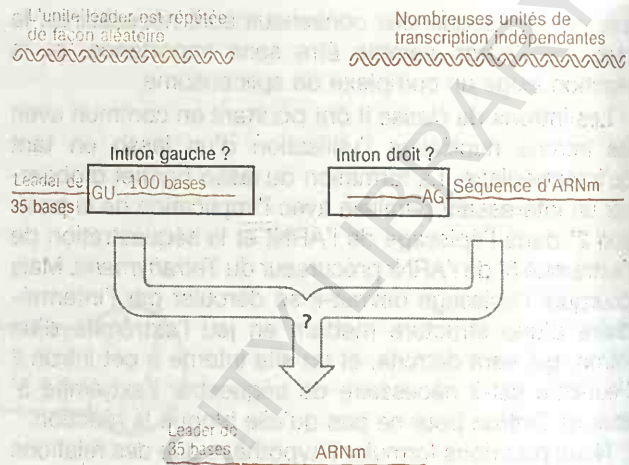


Figure 23.22

L'épissage en *trans* pourrait-il être responsable de la connexion du leader en 5' du trypanosome à des ARNm indépendants ?

jonction d'épissage droite juste avant la séquence présente dans l'ARNm mature. Le leader et l'ARNm pourraient-ils être connectés grâce à une réaction d'épissage en *trans* ? *

POUR EN SAVOIR PLUS

Nombre des premiers articles concernant l'épissage valent la peine d'être lus. Le système auto-catalytique *in vitro* de *Tetrahymena* a été découvert par Cech et al. (*Cell* 27, 487-496, 1981) et l'activité catalytique de l'ARN par Guerrier-Takada et al. (*Cell* 35, 849-857, 1983). Belfort et al. (*Cell* 41, 375-382, 1985) ont, les premiers, décrit l'épissage bactérien.

Dans la série « revues » l'ARN a été traité en tant que catalyseur par Altman (*Cell* 36, 227-229, 1984). Les réactions d'auto-épissage ont été répertoriées par Cech (*Int. Rev. Cytol.* 95, 3-22, 1985) et par Cech & Bass (*Ann. Rev. Biochem.* 55, 599-629, 1986). Padgett et al. (*Ann. Rev. Biochem.* 55, 1119-1150, 1986) ont passé en revue le mécanisme d'épissage. Plusieurs mini-revues ont été consacrées dans *Cell* à l'épissage. Les rapports entre les types d'épissage ont été étudiés par Cech (44, 207-210, 1986). Les caractéristiques particulières de l'épissage dans les organites ont été mises en relief par Lewin (22, 645-646, 1980). Lewin (22, 325-326, 1980) et Sharp (22, 643-646, 1981) ont abordé les mécanismes généraux impliquant des séquences consensus. Les structures et fonctions des snARN furent étudiées par Zieve (25, 296-297, 1981) et par Lerner & Steitz (25, 298-300, 1981) et les lassos analysés par Keller (39, 423-425, 1984).

* N. d. T. : L'hypothèse qu'un épissage en *trans* était à l'origine de la présence du « leader » en 5' des messagers de trypanosome est maintenant admise ; elle a, de plus, été étendue à d'autres systèmes d'eucaryotes inférieurs.

CHAPITRE 24

LE CONTRÔLE DE LA MATURATION DE L'ARN

Une des premières indications concernant l'écart de taille entre les gènes nucléaires et leurs produits chez les eucaryotes supérieurs a été fournie par l'étude des propriétés de l'ARN. Les gènes de structure codant pour des protéines sont transcrits dans le nucléoplasme. Mais l'ARN nucléoplasmique ne ressemble pas à l'ARNm. Sa taille est beaucoup plus grande, il est très instable, et sa complexité de séquence est beaucoup plus grande. Tirant son nom de sa large distribution en taille, il est appelé **ARN hétérogène nucléaire (hnARN)**.

Avant la découverte des gènes interrompus, il existait une vive controverse pour savoir si l'ARNm provenait de l'hnARN par une réduction de taille ou si l'hnARN avait une toute autre destination. Actuellement certaines des disparités constatées entre l'hnARN et l'ARNm peuvent être conciliées en invoquant la réduction de taille et de complexité qui se produit au cours de l'excision-épissage de l'ARN. De nombreux gènes de structure codant pour des protéines sont beaucoup plus longs que leurs ARNm et le transcrit primaire doit être au moins aussi long que le gène correspondant (il pourrait être plus long en incluant des séquences bordantes additionnelles).

Mais l'adjonction des introns aux exons n'explique pas entièrement la haute instabilité de l'hnARN. En moyenne, une petite proportion seulement des transcrits d'un gène produit effectivement un ARNm. Et certaines unités de

transcription sont transcrites dans le noyau, mais néanmoins ne semblent pas donner naissance à des ARNm cytoplasmiques. Ceci soulève la question de savoir si l'expression de certains gènes peut être contrôlée ou non au niveau de l'hnARN en déterminant si la voie de maturation produira de l'ARNm à partir d'un ARN nucléaire donné.

L'hnARN EST GRAND ET INSTABLE

La transcription des gènes dans le nucléoplasme représente seulement une petite partie de l'activité totale de synthèse de l'ARN de la cellule. La transcription intensive des gènes d'ARNr dans le nucléole est responsable de la plus grande partie (jusqu'à 90 %) de la synthèse de l'ARN. Il est difficile de suivre la synthèse globale de l'hnARN qui, en raison de sa faible proportion, est difficile à distinguer de l'ARNr. Cependant l'hnARN peut être analysé en étudiant seulement une partie de la population d'hnARN ou en utilisant des circonstances particulières.

Une des possibilités est d'examiner les fractions d'hnARN les plus grandes (qui sont distinctes des précurseurs d'ARNr car elles sédimentent au-delà de 45S). Il est également possible d'inhiber préférentiellement la synthèse d'ARNr par l'actinomycine. Une autre technique permet de suivre pendant un court moment le devenir

d'un marquage radioactif très bref (**pulse**) qui est incorporé dans l'hnARN plus rapidement que dans l'ARNr. Aucune de ces conditions n'est entièrement satisfaisante mais elles sont adéquates pour donner un aperçu des caractères généraux de la transcription nucléoplasmique.

Deux caractères saillants sont la taille et l'instabilité de l'hnARN. La plus grande partie du marquage radioactif incorporé dans l'hnARN est en fait recyclée dans le noyau. Elle n'entre même jamais dans le cytoplasme. La durée de vie de l'hnARN varie selon les différents types cellulaires mais est généralement comprise entre quelques minutes et environ une heure.

Le recyclage nucléaire démontre que *la plus grande partie de l'ARN est synthétisée et entièrement dégradée dans le noyau*. Ceci ne préjuge nullement du devenir des molécules considérées *individuellement* ou des parties des molécules. Ces données ne montrent pas si le matériel dégradé dans le noyau représente seulement des introns ou comprend des molécules entières d'hnARN.

Une des difficultés pour déterminer la taille exacte de l'hnARN est liée à sa tendance à s'agréger. Il existe dans l'hnARN une certaine proportion de structure secondaire

dont une grande partie représente des réactions entre des répétitions inversées d'une seule famille, la famille ALU ou son équivalent (voir Chapitre 30). Mais il existe aussi des réactions intermoléculaires ; celles-ci sont probablement responsables de la formation du matériel sédimentant très rapidement.

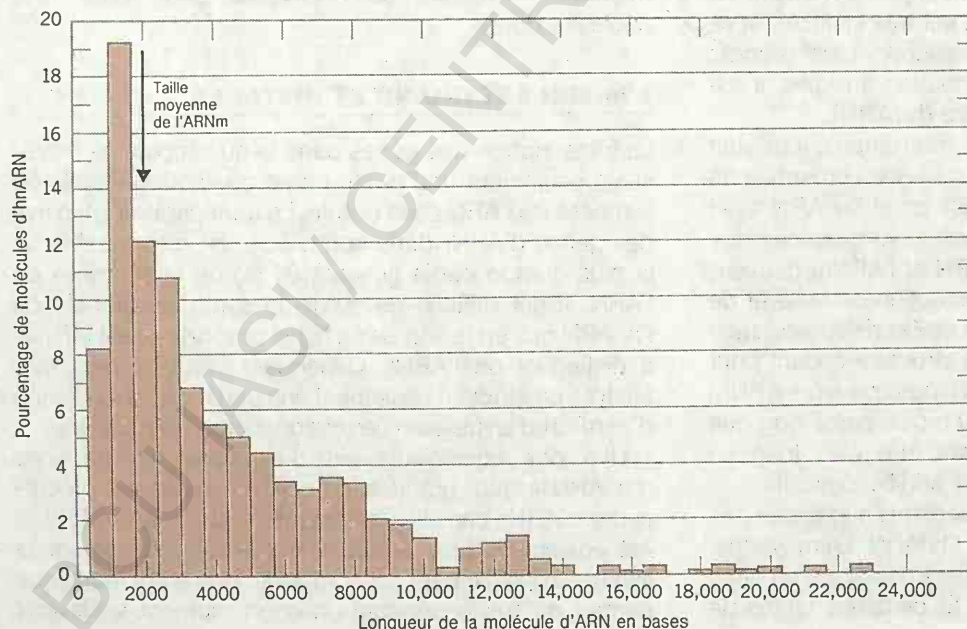
La distribution par taille de l'hnARN est comparée à la taille moyenne de l'ARNm dans la **Figure 24.1**. Cette distribution sous-estime la longueur des transcrits primaires car certaines des molécules les plus longues ont été probablement cassées par la méthode d'isolement. De plus, certaines des molécules pourraient être des intermédiaires partiellement dénaturés (se rappeler la série de précurseurs de l'ovomucoïde dans la Figure 23.15 qui sont tous présents dans la population d'hnARN à l'équilibre).

Mais malgré ces obstacles, il est clair que la majorité de l'hnARN est plus longue que l'ARN. Sur la base de l'analyse des « pulses » marqués incorporés dans les transcrits à l'état naissant, il semble maintenant que la distribution par taille des unités de transcription (chez les mammifères) a une moyenne de 8 000 à 10 000 pb. La distribution totale s'étend d'environ 2 000 à environ

Figure 24.1

L'hnARN a une très large distribution en taille allant de la taille moyenne de l'ARNm à plus de 14 000 bases.

L'histogramme mesure la longueur des molécules d'hnARN après étalement et microscopie électronique en conditions dénaturantes (qui devrait empêcher l'agrégation des molécules). Les molécules les plus courtes (inférieures à 1 500 bases) d'hnARN sont, probablement, principalement produites par des cassures au cours de la préparation.



14 000 pb. Ainsi la molécule moyenne d'hnARN est 4 à 5 fois plus grande que l'ARNm moyen, long de 1 800 à 2 000 bases.

L'étendue des longueurs suggère immédiatement que l'hnARN est trop long pour exister sous la forme d'une chaîne d'ARN déroulée à l'intérieur du noyau. En fait, on le trouve sous la forme de ribonucléoprotéines (hnRNP) où l'hnARN est associé à plusieurs protéines. Celles-ci semblent former un ensemble plus complexe que celles qui sont associées à l'ARNm. La masse de protéines dans l'hnRNP est d'environ 4 fois celle de l'hnARN.

La structure de l'hnRNP est obscure. La particule

ribonucléique commence à se former dès que l'ARN est transcrit ; il y a probablement peu d'ARN libre, même au début de la transcription. L'hnRNP peut se plier pour constituer une fibre qui, à son tour, se plie pour former une fibre plus épaisse ; il peut y avoir ainsi plusieurs niveaux de condensation.

L'hnRNP est associée à la **matrice nucléaire**, un réseau fibrillaire dense qui se trouve sur la face nucléaire de la membrane. Elle forme une sorte de paroi interne autour du noyau et s'étend aussi vers l'intérieur. La **Figure 24.2** est une photographie en microscopie électronique montrant la matrice nucléaire à l'intérieur des restes du cytoplasme. La structure des fibrilles n'est pas encore comprise, bien que nous sachions qu'elles sont composées d'un grand nombre de protéines.

La chromatine elle-même est attachée de manière intermittente à la matrice nucléaire, et il est probable que les transcrits primaires sont individuellement attachés peu après ou même pendant leur transcription. En fait leur maturation peut avoir lieu dans la matrice. Nous ne savons pas si cette situation est essentielle, c'est-à-dire si la maturation ne peut s'effectuer que là. Mais lorsque l'on pense aux événements impliqués dans la maturation de l'ARN, on doit se rappeler que le processus peut être topographiquement complexe et n'est pas simplement accompli par un ensemble d'enzymes flottant à l'intérieur du noyau.

Nous n'avons virtuellement aucune information sur la façon dont l'ARN mature est exporté vers le cytoplasme. Il est possible que sa migration puisse être liée spatialement à la maturation. De même, en ce qui concerne la voie d'exportation, il existe autour du noyau une série de pores qui pourraient fournir une voie de passage au transport de l'ARN mais nous ne savons pas s'ils sont utilisés effectivement dans ce but.

L'ARNm EST DÉRIVÉ DE L'hnARN

Les extrémités de certaines molécules d'hnARN portent les mêmes modifications que celles que l'on trouve dans l'ARNm. L'extrémité 5' acquiert une coiffe méthylée (partiellement), l'extrémité 3' est polyadénylée. Les similarités de ces modifications des extrémités suggèrent que l'hnARN et l'ARNm ont une relation de type précurseur-produit. [Également, parmi les protéines liées à l'hnARN, l'espèce majoritaire de 74 000 daltons associée au poly(A) pourrait être la même que celle présente dans l'ARNm.]

Le paradoxe qui est apparu lorsque l'on a constaté l'identité apparente des extrémités de l'hnARN et de l'ARNm était que l'hnARN est beaucoup plus grand que

Figure 24.2

La matrice nucléaire consiste en un réseau fibrillaire.

On a d'abord soumis une cellule HeLa à une extraction douce par un détergent non ionique pour enlever le matériel membranaire, puis on a utilisé un traitement à l'ADNase et haute concentration saline pour enlever la chromatine et une partie du cytosquelette (les fibres du cytoplasme environnant). Les petites structures en anneau à la surface de la matrice nucléaire pourraient être les pores nucléaires. Les grandes images sombres sont les nucléoles.

Photographie aimablement fournie par David Capco et Sheldon Penman.



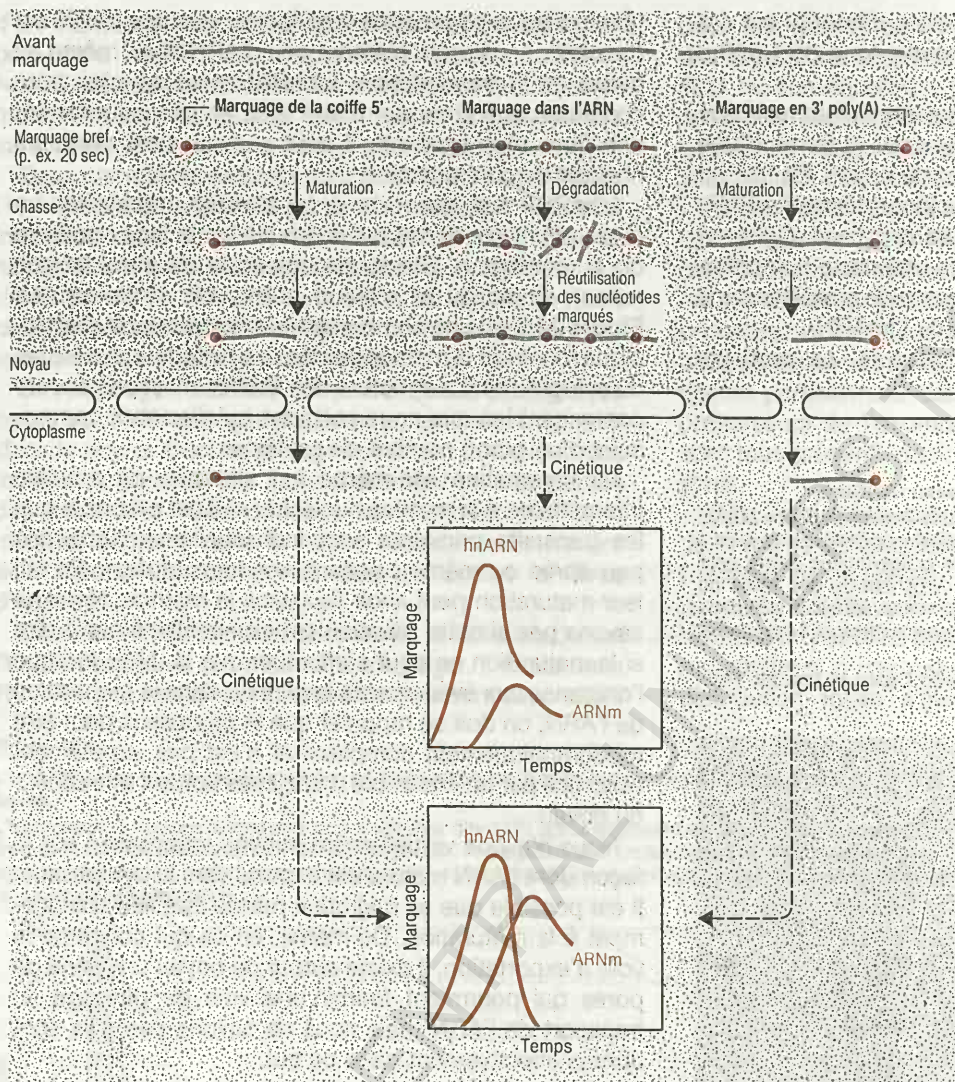


Figure 24.3

Le marquage radioactif des coiffes 5' ou des poly(A) 3' apparaît d'abord dans l'hnARN et est ensuite chassé dans l'ARNm, mais la plus grande partie du marquage du corps de l'hnARN est libérée par dégradation puis réutilisée.

l'ARNm. La façon dont les deux extrémités de la molécule se retrouvent dans l'ARNm n'apparaissait pas clairement. Maintenant nous comprenons que cette contradiction est résolue par la disparition des introns.

Des expériences de « **pulse-chase** » (marquage rapide suivi de chasse isotopique) peuvent être utilisées pour suivre le devenir des extrémités modifiées. Les résultats sont illustrés sur le diagramme de la **Figure 24.3**. Un précurseur marqué radioactivement est ajouté

pendant une très courte période de temps. C'est le « **pulse** » marqué. Il est ensuite ôté et remplacé par un précurseur non marqué. C'est l'épreuve de « **chasse** ». La quantité de marquage dans l'hnARN et l'ARNm est mesurée à différents intervalles de temps.

Pendant un temps court après le « **pulse** », le marquage s'accumule dans l'hnARN. Comme les molécules d'ARN sont instables, il décline ensuite rapidement. Des résultats différents sont obtenus selon que le marquage a été

incorporé dans l'une des extrémités ou à l'intérieur de l'ARN.

Comme les modifications des extrémités sont stables, elles peuvent être « chassées » de l'hnARN vers l'ARNm. En fait, la voie principale qu'empruntent les extrémités marquées pour quitter la population d'hnARN passe par la maturation de l'ARN, suivie de son départ pour le cytoplasme. Ainsi la baisse du marquage dans l'hnARN est accompagnée par une augmentation du marquage dans l'ARNm. La coïncidence de la baisse au niveau de l'hnARN et de l'accroissement dans l'ARNm démontre que les extrémités de l'ARNm sont dérivées des extrémités de l'hnARN. Pour ces parties de la molécule il existe une relation conventionnelle de type précurseur-produit entre l'hnARN et l'ARNm.

Il est probable qu'environ 20 % de l'hnARN sont « coiffés » à la suite de leur transcription. Une grande proportion des coiffes peut être chassée de l'hnARN vers l'ARNm. Bien que les données ne prouvent pas que toutes les coiffes 5' de l'hnARN donnent naissance à l'ARNm, cette relation semble plausible. L'adjonction de la coiffe survient très rapidement après le début de la transcription, comme en témoigne notre incapacité à caractériser les extrémités 5' triphosphate originelles sur les transcrits primaires.

À l'autre extrémité de la molécule, la longueur de poly(A) sur l'ARNm est légèrement plus courte que celle que l'on trouve dans l'hnARN. Quelques bases sont éliminées au cours du passage nucléocytoplasmique. On a quelque peu discuté pour savoir si tout le marquage du poly(A) peut être « chassé » du noyau vers le cytoplasme. La réponse peut varier selon les cellules et les circonstances, mais il est clair que la grande majorité au moins des queues poly(A) ajoutées à l'hnARN se retrouve dans l'ARNm.

Quand des nucléotides marqués sont incorporés dans les régions internes de l'hnARN, leur devenir principal est d'être libérés de la molécule quand celle-ci est dégradée. Puis ils sont réutilisés dans la synthèse d'ARN, entrant à la fois dans la constitution de l'ARNr et de l'hnARN. Le résultat est une diminution généralisée du marquage dans l'hnARN. Une petite proportion du marquage entre dans l'ARNm du cytoplasme mais en raison de sa situation minoritaire, l'apparition dans l'ARNm ne coïncide pas avec la chute dans l'hnARN, d'une manière qui démontrerait l'existence d'une relation de type précurseur-produit.

Les premières tentatives d'expériences de « pulse-chase » de nucléotides du corps de l'hnARN vers l'ARNm furent ainsi brouillées par le recyclage nucléaire de la majorité du marquage. Mais les améliorations de la

technique ont permis par la suite de montrer que l'ARNm est bien dérivé de l'hnARN. D'après de telles expériences, il est possible de calculer la proportion de la masse d'hnARN qui est convertie en ARNm. Celle-ci est d'environ 5 % dans les cellules de mammifères. Étant donné l'écart d'un facteur 5 dans la taille moyenne, cela correspond à une conversion d'environ 25 % des molécules d'hnARN.

Ceci veut dire que la réduction de la taille du transcrit liée à la perte des introns rend compte du devenir de seulement un quart environ de la population d'hnARN. Le reste consiste en des molécules qui sont *entièrement* recyclées dans le noyau : aucune partie de ces molécules n'est utilisée pour former de l'ARN cytoplasmique.

LA POLYADÉNYLATION ET LA FORMATION DES EXTRÉMITÉS 3'

Le poly(A) est ajouté aux extrémités de l'hnARN après la transcription, par une enzyme, la polyadénylate synthétase, qui utilise l'ATP comme substrat. L'addition de poly(A) à l'hnARN peut être empêchée par l'analogue 3' **déoxya-dénosine**, connue également sous le nom de **cordycépine**. Bien que la cordycépine n'arrête pas la transcription de l'hnARN, son addition empêche l'apparition de l'ARNm dans le cytoplasme. Ceci montre que la polyadénylation est *nécessaire* pour la maturation de l'hnARN en ARNm. [Ceci est vrai, bien sûr, seulement pour l'ARNm poly(A)⁺ ; l'ARNm poly(A)⁻ peut contourner cette exigence.]

Une partie seulement de l'hnARN est polyadénylée, environ 30 % dans les cellules de mammifères, ce qui peut être comparé à une valeur d'environ 70 % pour l'ARNm. Il y a souvent une corrélation grossière entre la proportion de molécules d'hnARN qui est polyadénylée et la proportion qui donne naissance aux molécules d'ARN poly(A)⁺ cytoplasmiques. Ceci suggère que la polyadénylation pourrait signaler qu'une molécule d'hnARN donnée doit subir la maturation. Cependant, bien que nécessaire, la polyadénylation n'est probablement pas suffisante par elle-même. En effet, dans ce cas, il y aurait toujours conservation du poly(A) entre le noyau et le cytoplasme.

L'extrémité 3' à laquelle le poly(A) est ajouté est engendrée par le clivage d'un ARN précurseur, ainsi que l'illustre la **Figure 24.4**. L'ARN polymérase transcrit au-delà du site correspondant au point où le poly(A) est ajouté, et des séquences dans l'ARN sont reconnues comme cibles d'une coupure endonucléolytique suivie par la polyadénylation. (Nous n'avons virtuellement aucune information sur le mécanisme de terminaison de l'ARN polymérase II, si bien que nous ne savons pas à

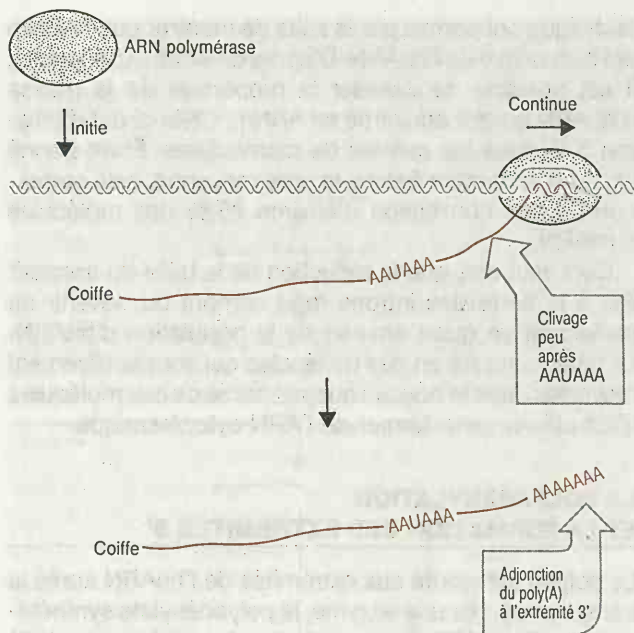


Figure 24.4

La séquence AAUAAA est nécessaire au clivage pour engendrer une extrémité 3' permettant la polyadénylation.

quoi ressemblerait un site de terminaison authentique. Il est possible que la terminaison soit « effilochée », survenant à de nombreux sites dans une région étendue, au lieu de résulter de la reconnaissance d'une séquence unique. *)

Une caractéristique des ARNm des eucaryotes supérieurs (mais pas chez la levure) est la présence d'une séquence AAUAAA dans la région de 11 à 30 nucléotides en amont du site d'addition du poly(A). Cette séquence est hautement conservée ; de manière seulement occasionnelle on peut trouver une différence sur une base.

Une délétion ou une mutation de l'hexamère AAUAAA empêche l'apparition de l'extrémité 3' habituellement polyadénylée. Cependant, chez un mutant ponctuel qui continue à être clivé avec une efficacité très réduite, les molécules qui sont clivées peuvent être polyadénylées normalement. La séquence AAUAAA fournit donc probablement un signal de clivage. Elle est nécessaire mais non suffisante en elle-même pour la réaction.

Le développement d'un système dans lequel la polyadénylation a lieu *in vitro* ouvre la voie permettant d'analyser ces événements. Dans un système qui fonctionne avec un précurseur purifié, l'addition d'un analogue de l'ATP qui ne peut pas être hydrolysé empêche la polyadé-

nylation. Cependant la réaction de clivage a lieu, ce qui montre que les deux réactions sont indépendantes.

Dans un système couplé qui fonctionne sur l'ARN synthétisé dans l'extrait, il existe un délai substantiel entre l'achèvement de la transcription et l'événement de polyadénylation. Ceci pourrait signifier que l'ARN ne peut former une structure appropriée de substrat, susceptible de maturation et de polyadénylation, *que* pendant la transcription, même si l'addition effective de poly(A) ne peut avoir lieu que plus tard.

La réaction *in vitro* est inhibée par les anticorps dirigés contre les snRNP, en particulier contre la snRNP U1. La snRNP U1 est-elle impliquée dans la polyadénylation ainsi que dans l'épissage ? Peut-être la snRNP U1 n'est-elle pas elle-même directement impliquée, mais fait partie d'une structure plus complexe qui est nécessaire à la réaction.

On a développé un système *in vitro* qui engendre d'authentiques extrémités 3' dans les ARNm d'histones (qui ne sont pas polyadénylés chez *X. laevis*). Quand un gène d'histone cloné (le gène H3 de l'oursin) est injecté dans l'ovocyte de *Xénope*, il est fidèlement initié et transcrit, mais la terminaison a lieu à des sites variables. Cependant quand un extrait nucléaire de l'oursin est injecté simultanément avec le gène, l'ARNm transcrit a une extrémité 3' correcte.

Le composant actif de l'extrait nucléaire est un facteur de 12S, d'une masse moléculaire d'environ $2,5 \times 10^5$. Il se comporte comme un complexe protéique quand il est fractionné. Cependant un petit ARN de l'oursin peut également produire d'authentiques extrémités 3', et cet ARN peut être isolé de la fraction 12S.

Le snARN U7 est long de 56 bases et n'est pas apparenté aux autres snARN. Il fonctionne probablement dans l'ovocyte de *Xénope* en raison de la présence d'un groupe de protéines de snRNP qu'il peut utiliser pour reformer une particule ribonucléoprotéique. L'extrémité 5' du snARN U7 est complémentaire de séquences courtes qui sont conservées de chaque côté du site de clivage 3' dans plusieurs gènes d'histones.

Nous pouvons conjecturer que la snRNP fonctionne par appariement de bases avec l'ARN cible, mais nous ne connaissons pas encore l'étape à laquelle la snRNP agit : il pourrait s'agir de la terminaison ou du clivage. Cependant l'implication d'une snRNP dans ce processus renforce largement le point de vue que beaucoup — et peut-être tous — les événements de maturation de l'hnARN se révéleront être dépendants d'interaction avec les snRNP.

* N. d. T. : Plusieurs régions d'ADN de mammifères au niveau desquelles l'ARN polymérase II cesse progressivement de transcrire les gènes ont été étudiées. Une caractéristique commune à ces séquences serait la richesse en T.

Le clivage et la polyadénylation précèdent habituellement l'épissage de l'ARN. Cependant, cette relation n'est pas causale en terme de mécanisme (bien que dans certains cas, où des jonctions d'épissage alternatives sont utilisées, le clivage de l'extrémité 3' puisse contrôler le choix de la jonction d'épissage : voir Chapitre 32).

Y A-T-IL UN CONTRÔLE APRÈS LA TRANSCRIPTION ?

La faible proportion de génome dans l'hnARN (inférieure à 5 % dans les cellules de mammifères) rend évident le fait que le contrôle de la transcription est un événement primaire de l'expression génique. Nous savons que les gènes exprimés abondamment — par exemple la globine ou l'ovalbumine — sont transcrits seulement dans les cellules dans lesquelles ils sont exprimés. Ce mode de contrôle est-il généralement applicable ? Quel est le degré de spécificité des séquences transcrites en hnARN dans chacun des types cellulaires ?

Étant donné la différence de taille entre l'hnARN et l'ARNm, nous nous attendons à une plus grande complexité de séquence pour l'hnARN. Pour les unités de transcription dans lesquelles un transcrit primaire est épissé pour produire un ARNm plus petit, la complexité de l'hnARN devrait inclure à la fois les exons et les introns, tandis que la complexité de séquence de l'ARNm devrait seulement inclure les exons.

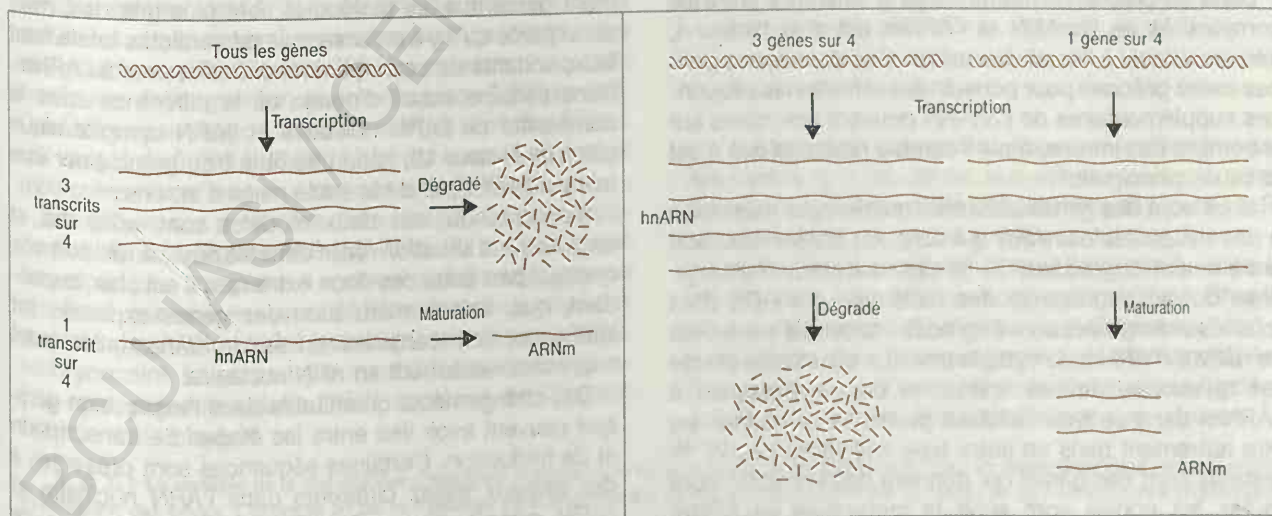
Malheureusement il est difficile de mesurer les complexités de séquence de l'hnARN avec beaucoup de précision. Premièrement, nous n'avons pas de tests fonctionnels pour isoler l'hnARN des autres ARN. Deuxièmement, l'utilisation des cinétiques d'hybridation est gênée par la présence dans l'hnARN de séquences répétitives qui sont entremêlées avec les séquences non répétitives qui forment la plus grande partie de l'ARNm.

La complexité de l'hnARN, déterminée à partir de sa capacité à saturer l'ADN non répétitif, atteint une valeur de 4 à 10 fois plus grande que celle de l'ARNm. Les expériences d'hybridations confirment que la population d'hnARN inclut bien toutes les séquences qui sont présentes dans l'ARNm cytoplasmique. Quelle est la nature des séquences additionnelles ?

Cette question peut être posée d'une autre façon. Les molécules d'hnARN, qui sont entièrement recyclées dans le noyau, représentent-elles des transcrits des *mêmes* gènes, ou bien de gènes *différents* de ceux qui donnent naissance à l'ARNm ? Les conséquences en sont illustrées dans la **Figure 24.5**.

Si les *mêmes* gènes sont impliqués dans une cellule de mammifère typique, un gène moyen est transcrit en molécules d'hnARN dont 75 % sont entièrement dégradées et 25 % fournissent 20 % de leur longueur (les exons) pour constituer de l'ARNm. Ainsi la maturation est extrêmement peu économe. Dans ce cas, la différence

Figure 24.5
Y a-t-il un gaspillage au cours de la maturation de tous les hnARN, comme le montre la Figure de gauche, ou bien certains transcrits subissent-ils une maturation efficace alors que d'autres ne donneraient pas naissance à l'ARNm, comme le montre la Figure de droite ?



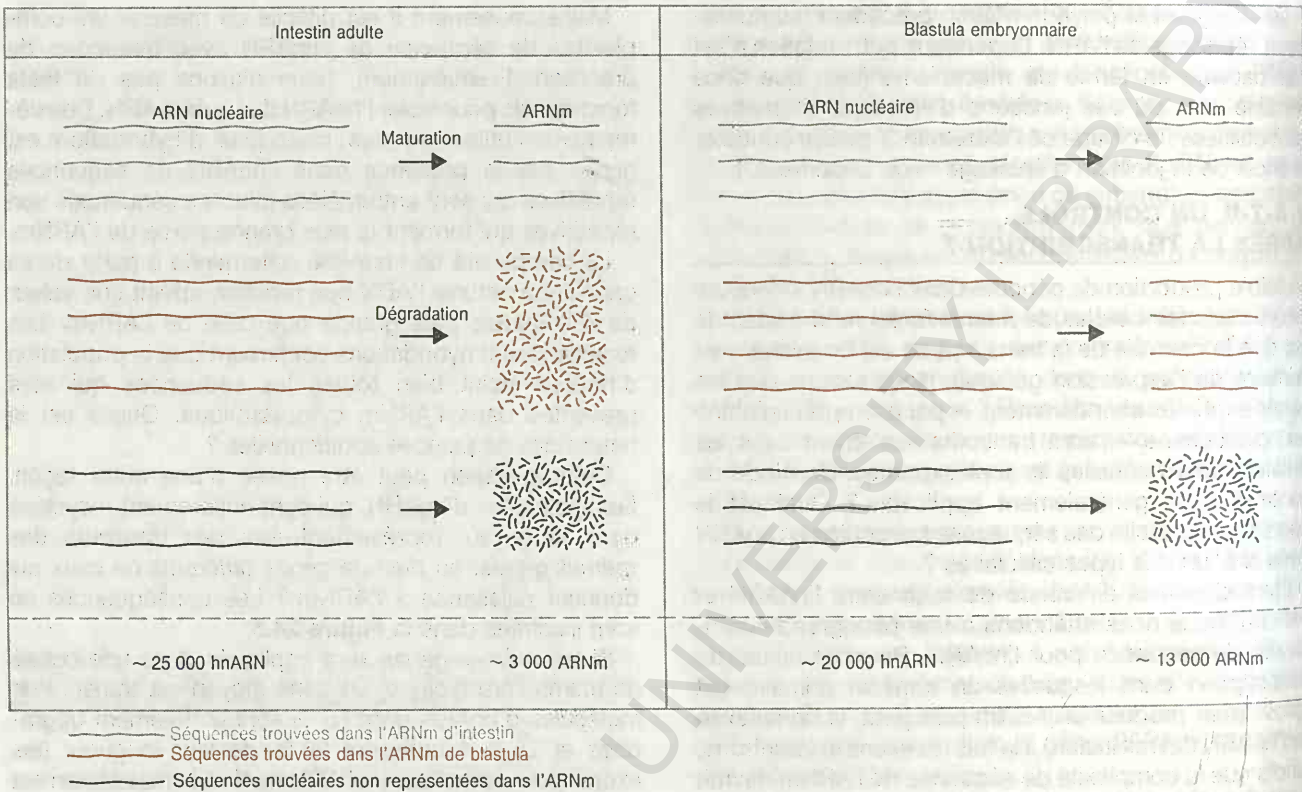


Figure 24.6

Les mêmes transcrits nucléaires sont dégradés dans l'intestin adulte, mais utilisés pour former de l'ARNm dans la blastula embryonnaire de l'oursin.

des complexités entre l'hnARN et l'ARNm est entièrement mise sur le compte de la perte des introns.

Dans les cellules de mammifères la différence entre les complexités de l'hnARN et l'ARNm est d'un facteur 4, très semblable à l'écart des tailles. Ces données ne sont pas assez précises pour pouvoir dire si *toutes* les séquences supplémentaires de l'hnARN peuvent être mises sur le compte des introns, mais il semble probable que c'est la cause principale.

Si ce sont des gènes *différents* qui donnent naissance à des molécules d'hnARN qui sont soit conservées, soit entièrement dégradées, 75 % des unités de transcriptions doivent représenter des molécules d'hnARN dont tous les exemplaires sont dégradés : aucune d'entre elles ne délivre d'ARNm au cytoplasme. (La conclusion stricte est qu'aucune de ces unités ne donne naissance à l'ARNm dans ce type cellulaire particulier, il pourrait en être autrement dans un autre type cellulaire.) Les 25 % restants sont des gènes qui donnent des transcrits dont toutes les copies vont subir la maturation en ARNm (conservant 20 % de leur longueur).

Dans ce cas, la complexité de l'hnARN devrait être beaucoup plus grande que celle de l'ARNm, non seulement parce que les molécules sont plus grandes, mais aussi parce qu'il y a un ensemble de transcrits totalement indépendants de ceux qui donnent naissance à l'ARNm. Dans certains tissus d'oursin où la différence entre la complexité de l'ARN nucléaire et l'ARN cytoplasmique est d'un facteur 10, l'écart semble trop grand pour être mis sur le compte de la seule perte d'introns.

Bien entendu, ces deux modèles sont extrêmes, et sans doute la situation réelle dans les noyaux réside-t-elle quelque part entre ces deux extrêmes. Il est clair, cependant, que, soit la maturation des gènes exprimés est inefficace, soit certaines unités de transcription sont exprimées seulement en ARN nucléaire.

Des changements quantitatifs dans l'expression génique peuvent avoir lieu entre les étapes de transcription et de traduction. Certaines séquences sont présentes à des niveaux assez différents dans l'ARN nucléaire et l'ARN cytoplasmique. Ceci signifie que le niveau de transcription n'est pas le seul facteur responsable de

l'établissement du niveau de l'ARNm. Des variations de stabilité ou d'efficacité de maturation pourraient intervenir pour modifier l'abondance relative d'une séquence d'ARN entre le noyau et le cytoplasme.

Une situation dans laquelle des variations quantitatives générales ont lieu est présentée par la transition que les cellules effectuent en culture, de l'état de repos (c'est-à-dire sans division) à l'état de croissance (quand elles progressent activement dans le cycle cellulaire).

Les cellules en croissance ont plus d'ARN que les cellules au repos, et en particulier elles ont une proportion accrue d'ARNm par rapport à l'ARNr. Mais il n'y a pas de changement dans le taux de transcription de l'hnARN, dans la proportion de molécules qui sont polyadénylées, ou dans la stabilité de l'ARNm. Ce qui se passe, c'est qu'une plus grande proportion de l'hnARN poly(A)⁺ est convertie en ARNm poly(A)⁺ dans les cellules en croissance.

La majorité des variations sont quantitatives plutôt que qualitatives. Certaines séquences nouvelles sont exprimées, mais, en majorité, la variation représente une augmentation de l'efficacité avec laquelle un ensemble constant de séquences présentes dans l'hnARN est converti en ARNm. Autrement dit, dans les cellules au repos, il y a une baisse dans l'efficacité de la conversion. A la lumière de notre question précédente sur la nature des séquences limitées au noyau, ceci indique que, au moins dans les cellules au repos, certaines d'entre elles correspondent bien à des copies en surplus de séquences exprimées.

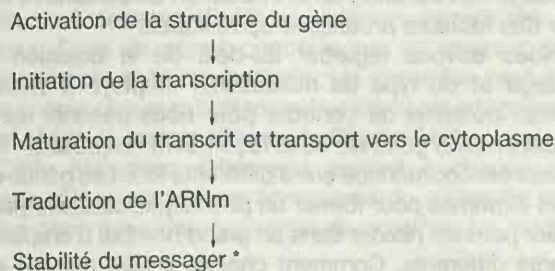
On trouve chez l'oursin une situation dans laquelle le contrôle de la sélection des séquences est exercé au niveau de l'ARN nucléaire. Chez *S. purpuratus*, les complexités totales de l'ARN nucléaire de plusieurs types cellulaires adultes et embryonnaires sont du même ordre de grandeur (175 000-225 000 kb consistant en des molécules de longueur moyenne d'environ 8 800 bases). Ceci n'a aucun rapport avec les complexités des ARNm qui varient de 6 000 à 26 000 kb et qui consistent en des molécules d'une longueur moyenne de 2 000 bases.

Les séquences présentes dans deux tissus sont comparées dans la Figure 24.6. L'ARNm de la blastula est beaucoup plus complexe. Il contient certaines séquences qui sont communes avec l'ARNm intestinal, mais la majorité des ARNm de la blastula est spécifique de l'état embryonnaire. La plupart (environ 80 %) de ces séquences sont présentes dans l'ARN nucléaire de l'intestin comme de la blastula.

Ceci signifie que le noyau intestinal contient des précurseurs de l'ARN qui sont confinés au noyau au lieu d'être utilisés pour produire de l'ARNm. Mais dans la blastula, ces mêmes séquences sont utilisées pour donner naissance à des ARNm. Ainsi, dans ce cas, de nombreux gènes sont transcrits indépendamment du fait qu'ils soient finalement exprimés ou non. Les transcrits sont sélectionnés pour subir une maturation et être transportés dans le cytoplasme seulement dans les tissus appropriés. Seule une minorité de gènes semble être contrôlée au niveau de la transcription.

MODÈLES DU CONTRÔLE DE L'EXPRESSION DES GÈNES

Il est trop tôt pour se faire une idée générale de la fréquence à laquelle les différents niveaux de contrôle sont utilisés pour réaliser l'étape à laquelle l'expression des gènes est contrôlée. Le concept de « niveau de contrôle » implique que l'expression du gène n'est pas forcément un processus qui se déroule automatiquement dès qu'il a commencé. Ce processus devrait en principe être contrôlé d'une manière spécifique à chaque gène, à chacune des différentes étapes. Nous pouvons distinguer (au moins) 4 points potentiels de contrôle formant la séquence :



L'existence de la première étape est impliquée par la découverte que les gènes eux-mêmes peuvent exister dans l'une ou l'autre de deux conditions structurales. Par rapport à l'état de la plus grande partie du génome, on trouve les gènes exprimés dans un état « actif » dans les cellules dans lesquelles ils sont exprimés (voir Chapitre 27). Le changement de structure est distinct de l'état de transcription et indique que le gène est « transcriptible ». Ceci suggère que l'acquisition de la structure « active » doit être la première étape de l'expression génique.

* N. d. T. : Le contrôle de la stabilité des messagers matures dans le cytoplasme est probablement aussi un niveau important de régulation de l'expression des gènes. L'exemple en est le messager de la vitellogénine de poulet dont la demi-vie de moins de 24 h en l'absence d'œstrogènes devient supérieure à 24 h en présence d'œstrogènes. La stabilité de transcrits d'oncogènes semble également souvent sujette à un contrôle extrêmement précis.

La transcription d'un gène dans l'état actif est contrôlée au stade de l'initiation, c'est-à-dire par l'interaction de l'ARN polymérase avec son promoteur. Ceci devient maintenant accessible à l'analyse dans les systèmes *in vitro* (voir Chapitre 9). Il n'y a actuellement pas de preuve d'un contrôle à des stades ultérieurs de la transcription dans les cellules eucaryotes, par exemple, par des mécanismes d'antitermination*.

Le transcrit primaire doit toujours être modifié par addition de coiffe et habituellement aussi par polyadénylation. Les introns doivent être excisés des transcrits des gènes interrompus. L'ARN mature doit être exporté du noyau du cytoplasme. Il n'est pas encore possible de distinguer ces étapes au niveau du contrôle, de sorte que la régulation de l'expression génique par la sélection des séquences au niveau de l'ARN nucléaire pourrait impliquer chacune ou toutes ces étapes. Nous n'avons aucune idée en termes moléculaires des événements qui peuvent être impliqués dans le contrôle de l'expression des gènes au niveau de l'ARN nucléaire, quantitativement ou qualitativement.

Enfin la traduction de l'ARNm dans le cytoplasme peut être contrôlée spécifiquement. Il y a peu d'arguments en faveur de l'emploi de ce mécanisme dans les cellules somatiques adultes, mais il a lieu dans certaines situations embryonnaires, comme cela est décrit dans le Chapitre 8. On suppose que ce mécanisme implique le blocage de l'initiation de la traduction de certains ARNm par des facteurs protéiques spécifiques.

Nous devons regarder au-delà de la question de l'usage et du type de mécanisme employé à chaque niveau potentiel de contrôle pour nous pencher sur la question plus générale de la façon dont l'expression des gènes est coordonnée entre différents loci. Les gènes qui sont exprimés pour former un phénotype cellulaire particulier peuvent résider dans un grand nombre d'emplacements différents. Comment chaque cellule active-t-elle la bonne combinaison de gènes pour produire un phénotype ?

On considère généralement que l'expression génétique des eucaryotes est sous contrôle positif : les gènes sont inactifs à moins qu'ils ne soient « ouverts ». Ceci est en accord avec la nécessité pour un gène de passer de l'état habituel inactif à la structure « active » avant de pouvoir être exprimé. Les stades ultérieurs du contrôle pourraient en principe être sujets à des contrôles positifs ou bien négatifs. Les modèles du contrôle de l'expression des

gènes eucaryotes supposent habituellement un contrôle positif mais peuvent facilement être représentés en terme de contrôle négatif.

Il y a deux types généraux de solution aux problèmes de l'expression de combinaisons chevauchantes de gènes dans différents phénotypes cellulaires. Les gènes pourraient être répétés, ou bien des gènes uniques pourraient être sujets à des contrôles multiples.

Un cas extrême de répétition de gènes serait d'avoir un exemplaire du gène pour chaque type cellulaire dans lequel il est exprimé. Chaque exemplaire serait sous un contrôle différent, spécifique d'un type cellulaire. Indépendamment du fait que les gènes spécifient des protéines identiques ou apparentées, il n'y a pas de problème lié à l'application d'une pression sélective si chaque gène est essentiel à un moment ou à un endroit particulier. (Quand différents gènes codent pour un ensemble de protéines remplissant la même fonction, la divergence entre les protéines pourrait accompagner l'évolution des différences de régulations, de façon à ce que chaque protéine soit adaptée au mieux aux circonstances de son utilisation.)

Il est certain que de nombreux gènes (et peut-être la plupart des gènes) sont présents à plus d'un exemplaire par génome. Il commence à apparaître qu'habituellement les exemplaires sont au moins légèrement différents. D'après les données limitées disponibles actuellement, il apparaît que les différents exemplaires ont tendance à être exprimés dans des situations cellulaires différentes.

Plutôt que de répéter l'ensemble du gène, un autre moyen d'engendrer la variation du produit est de ménager des exons alternatifs. Des changements dans le type d'épissage peuvent alors produire des protéines qui diffèrent seulement dans une région particulière de leur séquence. Le même mécanisme peut également être utilisé pour produire des ARNm dont les régions non traduites diffèrent alors que la protéine n'est pas modifiée**.

Quand il y a seulement une copie d'un gène, si son expression est requise dans plus d'un ensemble de circonstances, il existe la possibilité de le munir d'éléments régulateurs multiples, capables chacun de prendre en charge indépendamment son expression ; on trouve de tels arrangements dans les génomes bactériens et on peut également supposer leur existence chez les eucaryotes.

Un phénotype cellulaire est le résultat de l'expression

* N. d. T. : Il existe également un contrôle possible de l'élongation des transcrits : pose ou arrêt prématuré de transcription modulé en fonction des événements cellulaires, peut-être variation de la vitesse d'élongation. Des exemples de ces phénomènes ont été décrits au niveau de gènes de mammifères, notamment des oncogènes et du génome viral du virus HIV 1, responsable du SIDA.

** N. d. T. : De telles situations où il existe plusieurs transcrits d'un même gène, différents par leurs extrémités 5' non codantes, sont probablement le plus souvent secondaires à l'utilisation de promoteurs alternatifs, chacun initiant la transcription au niveau d'exons non codants 5' alternatifs.

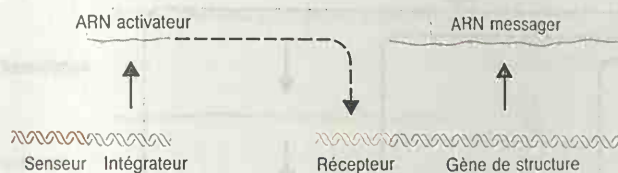


Figure 24.7

Les composants de base du contrôle des gènes eucaryotes pourraient comprendre un gène intégrateur, muni d'un site senseur, qui contrôle un gène de structure, muni d'un site récepteur.

coordonnée d'un ensemble particulier de gènes. Quand des gènes dispersés doivent être contrôlés de manière coordonnée, le même élément régulateur peut être répété à chaque locus.

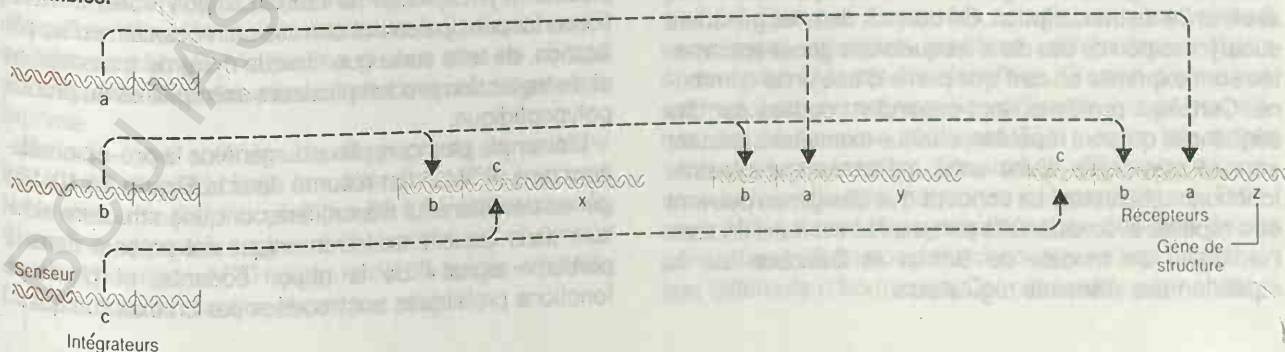
Des séquences d'ADN relativement courtes peuvent être utilisées comme sites de reconnaissance pour contrôler l'activité de gènes de structures contigus. Nous n'avons pas besoin de préciser comment ils contrôlent l'activité génique : leur action pourrait impliquer un changement structural avant la transcription, elle pourrait concerner l'initiation au niveau du promoteur, ou bien elle pourrait s'exercer sur le transcrit nucléaire.

Un modèle de réseau d'organisation permettant des contrôles multiples de chacun des gènes a été proposé par Britten et Davidson à partir du modèle de base illustré à la Figure 24.7. Un gène de structure est contrôlé positivement par un **site récepteur** adjacent. Celui-ci répond à une **molécule activatrice** considérée comme de l'ARN dans le modèle, mais qui pourrait alternatively être une protéine*. L'activateur est synthétisé par un **gène intégrateur** qui (bien sûr) est l'équivalent direct d'un gène régulateur bactérien. La seule différence est que le gène intégrateur est lui-même sujet au contrôle externe. L'élément de contrôle, qui lui est adjacent, est appelé **site senseur**. Il est supposé répondre aux fonc-

* N. d. T. : Et semble-t-il bien plus vraisemblablement.

Figure 24.8

Chaque gène de structure peut être activé par plus d'un récepteur et appartient donc à plusieurs ensembles.



tions qui contrôlent l'expression génique, les hormones, par exemple. Le modèle est exposé pour le contrôle avant la transcription mais il peut être facilement adapté aux autres niveaux.

Un seul activateur peut contrôler de nombreux gènes si chacun de ses loci-cibles possède un exemplaire du récepteur approprié. Il y a de nombreux précédents d'une telle répétition dans les loci bactériens qui sont sous un contrôle commun (voir Chapitre 10). Tous les gènes possédant le même récepteur composent un **ensemble** qui est exprimé en réponse à l'activateur. L'ensemble est assez équivalent à un opéron puisque tous les gènes sont exprimés ensemble. Il comprend donc vraisemblablement des gènes spécifiant des protéines de fonctions apparentées telles que les enzymes d'une voie métabolique.

Un gène de structure peut être exprimé dans des circonstances différentes s'il possède plusieurs récepteurs. Chaque récepteur est capable d'activer le gène en réponse à sa molécule activatrice. Si l'on place une série de récepteurs à côté d'un gène de structure, celui-ci peut être inclus dans plusieurs ensembles différents, comme le montre la Figure 24.8. Par une répétition appropriée des récepteurs, chacun des trois gènes de structures (x, y, z) appartient à deux des trois ensembles partiellement chevauchants (a, b, c).

Différents ensembles de gènes peuvent être contrôlés ensemble si l'on suppose qu'un senseur peut contrôler plusieurs intégrateurs. Ceci permet à un simple stimulus d'un seul senseur d'activer plusieurs ensembles de gènes. Tous les gènes contrôlés par un senseur sont désignés par le terme **batterie**. Un ensemble peut être inclus dans plusieurs batteries si les éléments intégrateurs sont répétés, comme l'illustre la Figure 24.9.

Cette hiérarchie d'éléments permet à chacun des gènes d'être exprimé en compagnie de n'importe quel autre gène avec une économie de contrôle. Le modèle

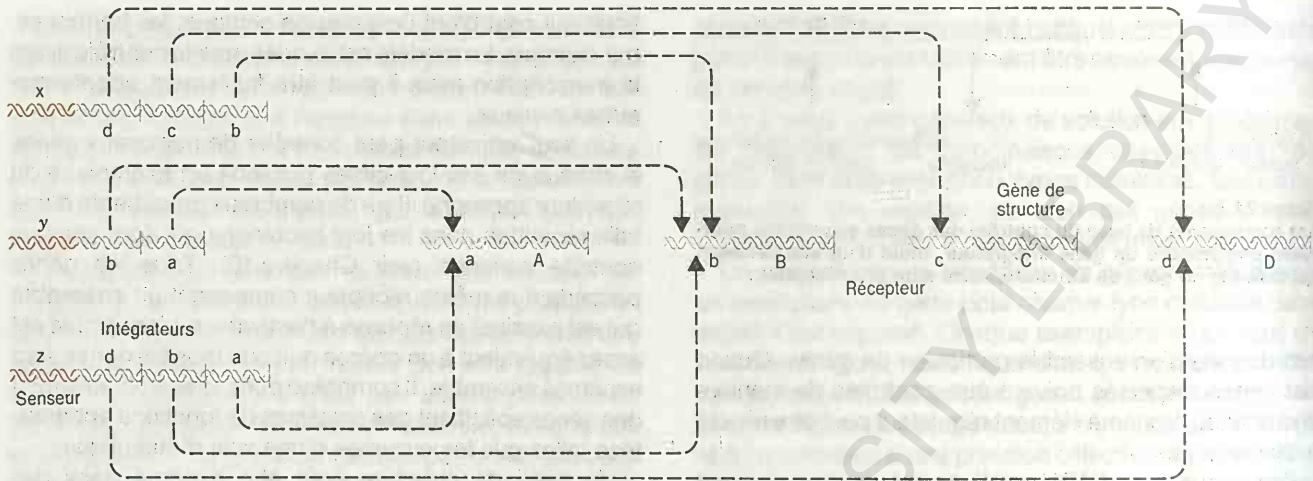


Figure 24.9

Les intégrateurs peuvent être répétés et chaque senseur peut contrôler plusieurs intégrateurs, de telle manière qu'un signal peut activer plusieurs ensembles de gènes, et que chaque gène peut être inclus dans plusieurs batteries.

nécessite que les séquences de l'intégrateur et du récepteur soient répétées, ce qui les définit donc comme membres de la composante répétitive de l'ADN. En effet, une motivation de la construction du modèle était de trouver un rôle aux séquences répétitives du génome.

Cependant, de manière ironique, il s'est avéré que les séquences répétitives, du moins telles qu'elles sont déterminées par la cinétique d'hybridation, n'ont probablement rien à voir avec la régulation. S'il y a des séquences répétitives régulatrices, elles doivent prendre probablement la forme de séquences consensus très courtes, beaucoup trop courtes pour être détectées dans l'analyse par hybridation.

LES PROPRIÉTÉS DES POLYPROTÉINES CELLULAIRES

L'exemple classique de l'opéron réalise le contrôle coordonné d'un groupe de gènes en les plaçant dans une seule unité de transcription. On connaît dans les génomes eucaryotes peu de cas dans lesquels des gènes apparentés sont exprimés en tant que partie d'une unité commune. Certaines protéines sont cependant codées par des séquences qui sont répétées, chaque exemplaire existant en tant que partie d'une unité qui comprend d'autres fonctions différentes. Le concept que des gènes peuvent être répétés avec différents partenaires est dans un sens l'antithèse du modèle de Britten et Davidson sur la répétition des éléments régulateurs.

Certaines des possibilités d'utilisation de séquences répétées sont représentées de manière frappante par la synthèse des **polyprotéines**. Le nom de polyprotéine a d'abord été utilisé pour décrire les produits multifonctionnels des génomes rétroviraux à ARN. Plusieurs protéines virales sont synthétisées sous la forme d'un précurseur commun, qui est clivé à des points spécifiques pour libérer des protéines distinctes fonctionnelles. La situation se conforme à la règle apparente qui demande que les systèmes de traduction eucaryotes n'utilisent que des messagers monocistroniques, mais elle permet à plusieurs protéines d'être synthétisées sous l'effet d'un seul événement de transcription.

Plusieurs formes de polyprotéines cellulaires ont été découvertes depuis lors. Dans certains cas, la polyprotéine contient une série de polypeptides identiques ou apparentés. Ainsi un précurseur d'énképhaline contient 6 exemplaires de Met-énképhaline (le peptide Tyr-Gly-Gly-Phe-Met), et également un exemplaire de Leu-énképhaline (Tyr-Gly-Gly-Phe-Leu). La polyprotéine fonctionne ici principalement comme un mécanisme d'amplification, de telle sorte que chaque cycle de transcription et de traduction produit plusieurs exemplaires du produit polypeptidique.

L'exemple plus complexe du gène de la pro-opiomélanocortine (POMC) est résumé dans la **Figure 24.10**. Les gènes de différents mammifères ont une structure similaire dans laquelle un intron unique est présent dans la portion « signal » de la région codante, et plusieurs fonctions protéiques sont codées par un exon unique.

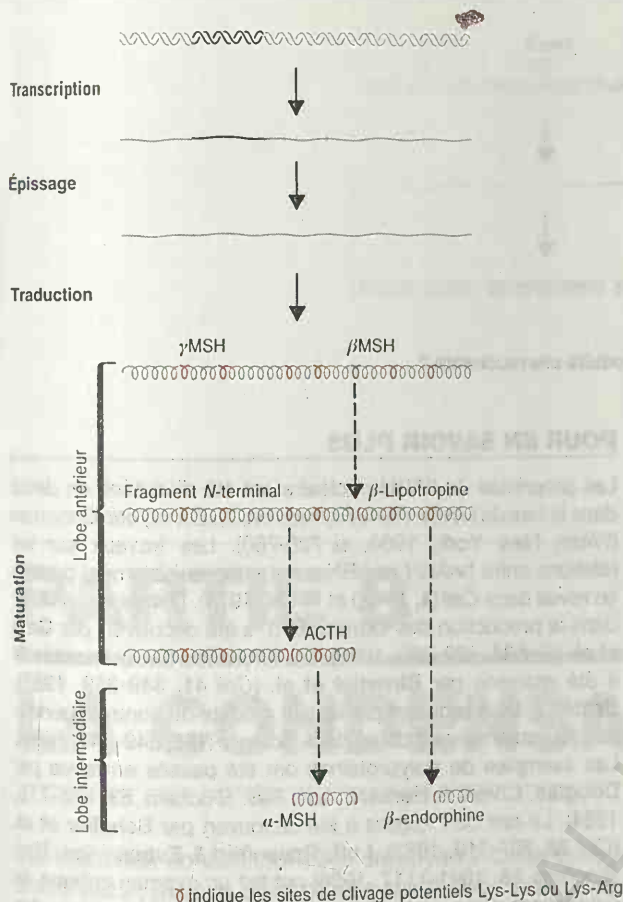


Figure 24.10

Le gène de la POMC est traduit en une polyprotéine dont les produits dépendent de la maturation spécifique de tissu.

Le produit polyprotéique unique de 32 000 daltons contient plusieurs hormones polypeptidiques connues et a également d'autres régions dont les séquences suggèrent qu'elles pourraient être aussi d'une certaine importance biologique. Tous les produits distincts, effectifs ou potentiels, sont limités par des dipeptides, consistant en des paires d'acides aminés basiques qui peuvent servir de cible au clivage. L'utilisation de ces sites de clivage est différente dans deux tissus dans lesquels le gène est exprimé.

Dans le lobe antérieur de l'hypophyse, la protéine POMC est d'abord clivée une fois, libérant le fragment N-terminal et la β-lipotropine (les 70 à 90 acides aminés C-terminaux) puis le fragment N-terminal est clivé pour donner l'hormone ACTH (longue de 39 acides aminés). Dans ce tissu la réaction ne va pas plus loin.

Dans le lobe intermédiaire, l'ACTH est clivée de nouveau, libérant l'α-mélanotropine (α-MSH) qui se compose des 13 acides aminés N-terminaux. La β-lipotropine est également clivée une fois, libérant la β-endorphine, analgésique, qui se compose des 31 acides aminés C-terminaux.

En plus de ces hormones connues, il existe deux autres séquences ayant une activité MSH potentielle ; elles sont dénotées β-MSH et δ-MSH mais nous ne savons pas encore quand et si elles sont produites.

Quel est le but de cette forme d'organisation ? Nous ne comprenons toujours pas l'ensemble des fonctions de ces hormones bien que nous sachions qu'elles sont produites en réponse à un stress douloureux. La fonction des chevauchements de séquences entre les hormones alternatives ne deviendra claire que lorsque nous comprendrons mieux leurs activités biologiques.

Ces cas permettent d'établir le principe qu'un gène unique peut donner naissance à des produits différents (quoique chevauchants) en vertu de variations dans la voie de maturation des protéines (régulée probablement par les enzymes de maturation des tissus correspondants). Une autre possibilité est que des exemplaires répétés des gènes de polyprotéines pourraient permettre à la même protéine d'être synthétisée en compagnie de partenaires différents. Le principe est illustré dans la Figure 24.11.

Deux gènes de polyprotéines contiennent les séquences des différentes protéines A,B,C, et A,B,D. Ainsi l'activation de chacun des gènes conduit à la synthèse de A et de B mais dans un cas, celle-ci est accompagnée par la production de C, dans l'autre cas, par la production de D. Cette situation est formellement équivalente à la répétition de gènes de structure dans des unités de transcription polycistroniques, bien que les produits soient séparés au stade de la maturation des protéines plutôt qu'à celui de la traduction.

Une famille de gènes codant pour l'ELH, un neuropeptide du mollusque marin *Aplysia*, pourrait correspondre à cette catégorie. L'ADN d'*Aplysia* contient environ 5 gènes de l'hormone ELH (un polypeptide de 36 acides aminés). Bien que l'ensemble complet de gènes reste à caractériser, plusieurs espèces différentes d'ARNm codant pour ELH ont été identifiées. Chacun d'entre eux est traduit en une polyprotéine différente. Les polyprotéines ont toutes une séquence ELH ; mais les autres séquences représentent des produits différents qui peuvent inclure d'autres hormones polypeptidiques. Différents ARN peuvent être synthétisés dans différents tissus, permettant la libération dans chaque cas d'une combinaison différente d'hormones. Ce mécanisme permet donc

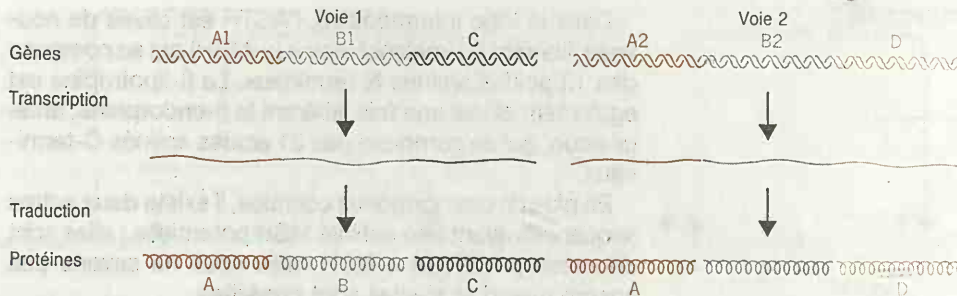


Figure 24.11

Y a-t-il des gènes codant pour des polyprotéines consistant en des produits chevauchants ?

à des combinaisons chevauchantes d'hormones d'être synthétisées, grâce à l'activation des membres appropriés d'une famille de gènes répétés.

Il est notable que ces exemples de polyprotéines cellulaires impliquent la synthèse de petits produits (généralement de moins de 40 acides aminés chacun). Ceci reflète-t-il quelque difficulté à synthétiser des petites protéines de *novo*, ce qui a conduit à la nécessité d'un clivage à partir de précurseurs plus longs (ouvrant ainsi la voie à l'évolution de produits alternatifs) ? Il sera intéressant de caractériser les gènes qui codent pour plusieurs polyprotéines apparentées. Se révéleront-ils être assemblés par brassage des exons, de la manière décrite dans le Chapitre 18 ou bien l'ensemble du processus a-t-il évolué à un stade ultérieur de l'expression de gènes répétés dont les ancêtres n'avaient pas de fonctions multiples ?

POUR EN SAVOIR PLUS

Les propriétés de l'ARN nucléaire ont été discutées en détail dans le livre de **Lewin** *Gene Expression, 2., Eucaryotic Genomes* (Wiley, New York, 1980, p. 728-760). Les travaux sur les relations entre hnARN et ARNm ont été préalablement passés en revue dans *Cell* (4, 11-20 et 77-94, 1975). Le rôle des snRNP dans la production des extrémités 3' a été découvert par **Galli et al.** (*Cell* 34, 822-828, 1983). La maturation des extrémités 3' a été analysée par **Birnstiel et al.** (*Cell* 41, 349-359, 1985). **Britten & Davidson** ont publié leur modèle du contrôle génétique (qui commence à dater) dans *Science* (165, 349-347, 1969). Les exemples de polyprotéines ont été passés en revue par **Douglas, Civelli & Herbert** (*Ann. Rev. Biochem.* 53, 665-715, 1984). Le cas de l'*Aplysia* a été découvert par **Scheller et al.** (*Cell* 28, 707-719, 1982). **Leff, Rosenfeld & Evans** (*Ann. Rev. Biochem.* 55, 1091-1117, 1986) ont fait un examen critique de l'utilisation d'une maturation alternative pour engendrer des produits multiples à partir d'une seule unité de transcription.

PARTIE 8

LA CONDENSATION DE L'ADN

Les bactéries contiennent-elles un noyau ? Cette question a remué de nombreux chercheurs ; mais, à présent, elle paraît avoir aussi peu de sens que les éternelles controverses sur la paternité des pièces de Shakespeare, qui ont été finalement résolues par la considération qu'un autre individu du même nom devait en être l'auteur. La propriété importante du noyau des cellules animales et végétales n'est pas tant dans son apparence ou sa forme que dans les changements réguliers, reliés à des taches biologiques définies, qu'on peut le voir subir.

Erwin Chargaff, 1947

CHAPITRE 25

AU SUJET DES GÉNOMES ET DES CHROMOSOMES

Un principe général est évident dans l'organisation de tout le matériel génétique cellulaire. Il existe sous la forme d'une masse compacte dans une zone délimitée ; et ses différentes activités, telles que la réplication et la transcription, doivent être accomplies à l'intérieur de ces limites. Les principales questions qui restent sans réponse concernent l'organisation fondamentale de ce matériel, de même que les changements qui surviennent quand il devient actif.

L'état condensé de l'acide nucléique dépend de la présence de protéines spécifiques auxquelles il se lie. Ces protéines sont basiques et leurs charges positives neutralisent les charges négatives de l'acide nucléique. Les relations des protéines avec l'ADN (ou l'ARN) déterminent sa structure.

Un problème commun est présenté par la condensation de l'ADN dans les phages et les virus, ou dans les cellules bactériennes et les noyaux eucaryotes. La longueur de l'ADN sous une forme moléculaire déroulée excéderait largement les dimensions du compartiment qui le contient. L'ADN (ou, dans le cas de certains virus, l'ARN) doit donc être extrêmement comprimé pour s'adapter à l'espace disponible.

La grandeur de l'écart entre la longueur de l'acide nucléique et la taille de son compartiment est évidente sur les exemples résumés dans le **Tableau 25.1**. Pour

les bactériophages et les virus eucaryotes, qu'ils soient longs et minces (filamenteux), ou approximativement sphériques (icosaédriques), une grande quantité d'acide nucléique, ADN ou bien ARN, simple brin ou double brin remplit en effet le compartiment.

Pour les bactéries ou les compartiments des cellules eucaryotes, l'écart est encore plus grand. L'ADN est contenu dans une région compacte qui occupe seulement une partie du compartiment. On voit le matériel génétique sous la forme du **nucléoïde** chez les bactéries, et d'amas de **chromatine** dans les noyaux eucaryotes à l'interphase (entre les divisions). Parmi les exemples donnés dans le tableau, le problème atteint son point culminant chez l'homme, avec la nécessité d'adapter 1,8 mètre d'ADN humain à un noyau dont le diamètre peut être aussi petit que $6 \mu\text{m}$ ($6 \times 10^{-6} \text{ m}$).

La condensation de la chromatine est souple. Elle change au cours du cycle cellulaire eucaryote. Au moment de la division (mitose ou méiose) le matériel génétique devient encore plus compact et des **chromosomes** mitotiques distincts deviennent identifiables.

La compression globale de l'ADN peut être décrite par le **rapport de condensation**, c'est-à-dire la longueur de l'ADN divisée par la longueur de l'unité qui le contient. Par exemple, le plus petit chromosome humain contient $4,6 \times 10^7$ pb d'ADN (environ 10 fois la taille du génome

Tableau 25.1

La longueur de l'acide nucléique est beaucoup plus grande que les dimensions du compartiment qui le contient.

Compartiment	Forme	Dimensions	Type d'acide nucléique	Longueur	
TMV	filament	$0,008 \times 0,3 \mu\text{m}$	1 ARN simple-brin	$2 \mu\text{m}$	6,4 kb
Phage fd	filament	$0,006 \times 0,85 \mu\text{m}$	1 ADN simple-brin	$2 \mu\text{m}$	6 kb
Adénovirus	icosaèdre	$0,07 \mu\text{m}$ de diamètre	1 ADN double-brin	$11 \mu\text{m}$	35 kb
Phage T4	icosaèdre	$0,065 \times 0,10 \mu\text{m}$	1 ADN double-brin	$55 \mu\text{m}$	170 kb
<i>E. coli</i>	cylindre	$1,7 \times 0,65 \mu\text{m}$	1 ADN double-brin	1,3 mm	$4,2 \times 10^3$ kb
Mitochondrie (homme)	sphéroïde allongé	$3,0 \times 0,5 \mu\text{m}$	~ 10 ADN double-brins identiques	$50 \mu\text{m}$ $5 \mu\text{m}$ chacun	16 kb
Noyau (homme)	sphéroïde	$0,06 \mu\text{m}$ de diamètre	46 chromosomes d'ADN double-brin	1,8 m	6×10^6 kb

de la bactérie *E. coli*). Ceci est équivalent à $14\,000 \mu\text{m} = 1,4 \text{ cm}$ d'ADN déroulé. Au moment de condensation maximum au cours de la mitose, le chromosome est long d'environ $2 \mu\text{m}$. Ainsi, le rapport de condensation de l'ADN dans le chromosome peut atteindre 7 000.

Les rapports de condensation ne peuvent être établis avec cette certitude pour les structures globales plus amorphes des nucléoïdes bactériens ou de la chromatine eucaryote. Cependant on estime que les chromosomes mitotiques sont probablement 5 à 10 fois plus compacts que la chromatine à l'interphase, qui a donc un rapport de condensation typique de 1 000 à 2 000.

Une des questions majeures restant sans réponse concerne la *spécificité* de la condensation. L'ADN est-il replié selon un mode *particulier*, ou bien est-il différent dans chacun des exemplaires du génome ? Comment change le mode de condensation quand un segment d'ADN est répliqué ou transcrit ?

LA CONDENSATION DES GÉNOMES VIRAUX DANS LEUR REVÊTEMENT

Dans la perspective de la condensation de séquences *distinctes*, il existe une différence importante entre un génome cellulaire et un virus. Le génome cellulaire est essentiellement de taille indéfinie ; le nombre et l'emplacement de séquences distinctes peut être changé par duplication, délétion et réarrangement. Ainsi la condensation de son ADN requiert une méthode *générale*, insensible au contenu total ou à la distribution des séquences. Au contraire, deux restrictions déterminent les exigences

d'un virus. La quantité d'acide nucléique à emballer est *prédéterminée* par la taille du génome. Et il doit s'adapter entièrement à l'intérieur d'un revêtement assemblé à partir d'une protéine ou de protéines codées par les gènes viraux.

Une particule virale est trompeusement simple dans son apparence superficielle. Le génome d'acide nucléique est contenu à l'intérieur d'une **capside**, une structure symétrique ou quasi symétrique assemblée à partir d'une ou de quelques protéines seulement. Attachées à la capsid ou incorporées à l'intérieur de celle-ci, peuvent exister d'autres structures assemblées à partir de protéines distinctes et nécessaires pour l'infection de la cellule hôte.

Vues sous cet angle, les limites pour construire la particule virale sont plutôt étroites. Le volume interne de la capsid est rarement beaucoup plus grand que le volume de l'acide nucléique qu'il doit contenir. La différence est habituellement inférieure à un facteur deux et souvent le volume interne est à peine plus grand que l'acide nucléique.

Dans sa forme la plus extrême, la limitation qui fait que la capsid doit être assemblée à partir des protéines codées par le virus signifie que la coque entière est construite à partir d'un seul type de sous-unité. Les règles de l'assemblage de sous-unités identiques en structure fermée restreignent la forme de la capsid à l'un ou l'autre des deux types suivants. Les sous-unités protéiques peuvent s'empiler séquentiellement en un arrangement hélicoïdal pour former une structure **filamenteuse** ou en bâtonnet. Ou bien, elles peuvent former une coque pseudo-sphérique ; ce type de structure est en fait un polyèdre ayant une **symétrie icosaédrique**. Certaines

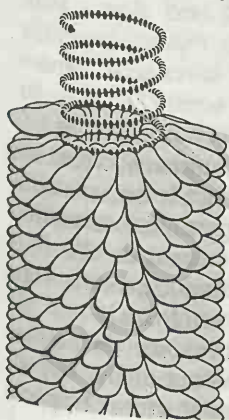
capsides virales sont assemblées à partir de plusieurs types de sous-unités protéiques, mais bien que ceci puisse augmenter la diversité des structures qui peuvent être formées, les capsides virales se conforment cependant toutes aux classes générales des filaments quasi cristallins ou des icosaèdres.

Il y a deux types de solutions au problème de la construction d'une capside qui contient un acide nucléique. La coque protéique peut être assemblée autour de l'acide nucléique en condensant l'ADN ou l'ARN par des interactions protéines-acides nucléiques au cours du processus d'assemblage, ou bien la capside peut être construite à partir de ses constituants sous la forme d'une coque vide, après quoi l'acide nucléique doit être inséré, étant condensé au fur et à mesure qu'il entre.

L'assemblage de la capside autour du génome a lieu dans le cas des virus à ARN simple brin. L'exemple le mieux caractérisé est celui du TMV (virus de la mosaïque du tabac). L'assemblage commence au niveau d'une structure en épingle à cheveux double brin qui réside dans la séquence d'ARN. À partir de ce **centre de nucléation**, il progresse dans les deux directions le long de l'ARN jusqu'à ce qu'il atteigne les extrémités. L'unité de la capside est un disque composé de deux couches, chaque couche contenant 17 sous-unités protéiques identiques. Le disque forme une structure circulaire, qui est convertie en forme hélicoïdale par interaction avec l'ARN. L'ARN est enroulé de façon hélicoïdale sur la partie interne de la coque protéique comme l'illustre la **Figure 25.1**.

La longueur complète de ce génome d'ARN tient dans une structure déterminée par son interaction avec la coque protéique. Un arrangement analogue est présenté par les capsides sphériques qui contiennent un ARN

Figure 25.1
Un passage en hélice pour l'ARN du TMV est créé par l'empilement des sous-unités protéiques du virion.



simple brin, comme, par exemple, dans le cas du virus de la mosaïque jaune du navet (TYMV). Le principe commun est que *la position de l'ARN à l'intérieur de la capside est déterminée directement par sa liaison aux protéines de la coque*.

Les capsides sphériques des virus à ADN sont assemblées d'une manière différente qui est bien caractérisée chez les phages lambda et T4. Dans chaque cas, une tête vide est assemblée à partir d'un petit ensemble de protéines. *Puis le génome double brin est inséré dans la tête*, ce processus étant accompagné par un changement de structure dans la capside.

La **Figure 25.2** résume l'assemblage de lambda. Il commence avec une petite tête qui contient une protéine « cœur ». Celle-ci est convertie en une tête vide de forme distincte. Puis l'empaquetage de l'ADN commence, la tête augmente de taille, tout en restant de la même forme, et finalement, la tête remplie est scellée par l'addition de la queue.

Pour lambda comme pour T4, l'ADN qui doit être inséré dans la tête vide prend la forme de molécules **concatémériques**. Celles-ci sont constituées de génomes multiples joints bout à bout comme le représente la **Figure 25.3**. Chaque phage a son propre mécanisme pour reconnaître la quantité correcte d'ADN qu'il doit insérer.

Les extrémités du génome de lambda sont marquées par des séquences appelées sites *cos*. Le clivage a lieu au site *cos* de gauche (tel qu'il est défini sur la carte usuelle) pour produire une extrémité libre qui est insérée dans la capside. L'insertion de l'ADN continue jusqu'à ce que le site *cos* de droite soit rencontré ; il est alors clivé pour produire l'autre extrémité. L'extrémité qui entre la dernière dans la capside au cours de l'assemblage sort la première quand une nouvelle cellule hôte est infectée.

N'importe quel ADN contenu entre les deux sites *cos* peut être encapsidé. (Ceci est à la base de la technique de clonage dans les « cosmides » qui est décrite dans le Chapitre 16.) La distance entre les sites *cos* ne peut être modifiée que légèrement par rapport à la taille habituelle de l'ADN de lambda. L'encapsidation n'a pas lieu si la distance est trop grande ou bien trop petite. Ceci démontre qu'il doit y avoir *suffisamment* d'ADN pour achever la réaction d'encapsidation, tout en montrant qu'il n'y a de place dans la tête que pour une très petite quantité d'ADN supplémentaire (environ 15 %).

Dans le phage T4, l'insertion commence à un endroit *au hasard* dans le précurseur concatémérique. Elle continue jusqu'à ce que l'équivalent d'un génome d'ADN (de quoi remplir la tête) ait été inséré. Ceci implique l'existence d'un mécanisme permettant de mesurer la quantité d'ADN. [En fait, la quantité insérée est légèrement plus

La pro-tête I
contient
la protéine « cœur »

La pro-tête II
est vide

L'encapsidation
commence

La particule grise
est partiellement
remplie et a une tête
augmentée de volume

Tête pleine
(particule noire)

La tête est stabilisée
pour l'attachement
de la queue

Particule
mature

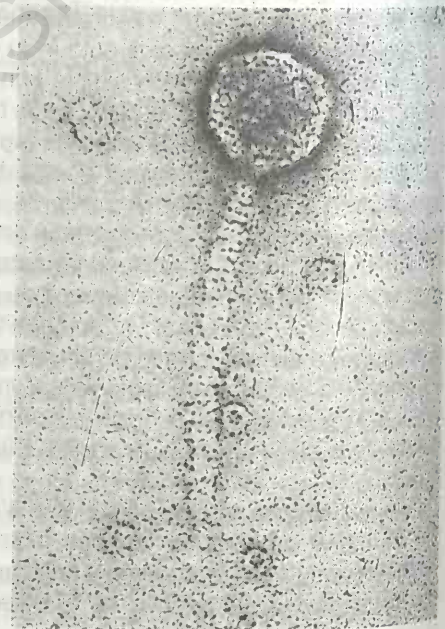
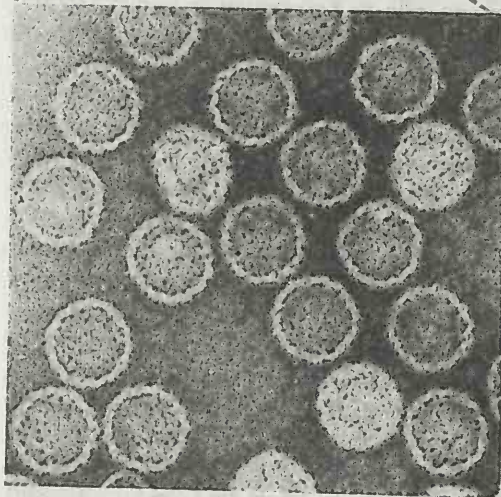
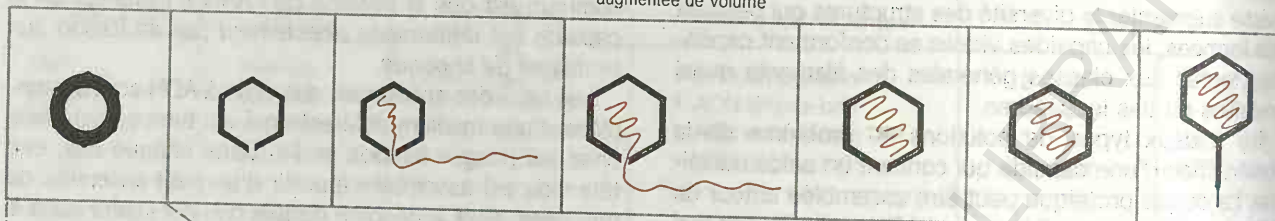


Figure 25.2

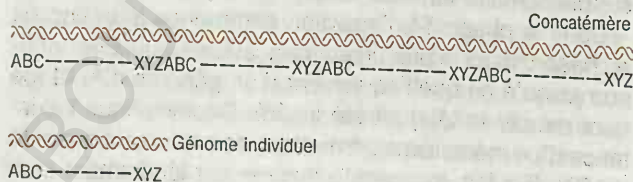
La tête vide du phage lambda change de forme et augmente de volume lorsqu'elle est remplie par l'ADN.

Les photographies en microscopie électronique montrent les particules au début et à la fin de la voie de maturation.

Photographies aimablement fournies par A. F. Howatson.

Figure 25.3

L'ADN concatémère consiste en une série de génomes de phages en tandem.



grande que la longueur de l'unité de génome, créant une **redondance terminale** correspondant à la longueur additionnelle. Dans les termes de la Figure 25.3, le premier virion contient l'ADN de A à A, le suivant de B à B, et ainsi de suite de telle manière que chaque génome a une lettre (différente) répétée à chaque extrémité.]

Or un ADN double brin considéré sur de courtes distances forme une tige passablement rigide. Cependant, il doit être comprimé dans une structure compacte

pour s'adapter à l'intérieur de la capsid. Nous aimerions savoir si l'encapsidation implique un enroulement souple de l'ADN dans la tête ou nécessite des flexions abruptes.

On connaît peu de choses sur le mécanisme de l'encapsidation en dehors du fait que la capsid contient des protéines internes en plus de l'ADN. Une possibilité est qu'elles puissent former une sorte d'armature sur laquelle l'ADN se condense. (Ceci serait une contrepartie de l'utilisation des protéines de la coque dans les virus de plantes mentionnés précédemment.)

Quelle est la spécificité de cette encapsidation ? Elle ne peut pas dépendre de séquences particulières, puisque des délétions, insertions et substitutions ne sont pas capables d'interférer avec le processus d'assemblage. Les relations entre l'ADN et la tête ont été examinées directement en déterminant quelles régions de l'ADN peuvent être pontées chimiquement aux protéines de la capsid. La réponse surprenante est que toutes les régions de l'ADN sont plus ou moins également susceptibles d'être pontées. Ceci veut probablement dire que quand l'ADN est inséré dans la tête il suit une règle générale pour se condenser, mais le mode de condensation n'est pas déterminé par des séquences particulières.

Ces mécanismes variés d'assemblage permettent tous d'accomplir la même finalité : l'emballage d'une molécule unique d'ADN ou d'ARN dans la capsid. Cependant, certains virus ont des génomes qui comprennent de multiples molécules d'acide nucléique. Le réovirus contient 10 segments d'ARN double brin qui doivent tous être emballés dans la capsid. On ne connaît rien sur la manière dont le processus d'assemblage sélectionne un exemplaire de chaque molécule différente afin de réunir un assortiment complet d'information génétique.

Certains virus de plantes sont multipartites (à divisions multiples) : leur génome consiste en des segments dont chacun est emballé dans une capsid *différente*. Un exemple en est le virus de la mosaïque alfalfa, qui a quatre ARN simple brin différents, chacun étant emballé indépendamment dans un revêtement comprenant la même sous-unité protéique. Une infection réussie dépend de l'entrée des différents types dans la cellule.

Les quatre constituants du virus existent sous la forme de particules de différentes tailles. Ceci veut dire que la même protéine de capsid peut emballer chacun des ARN dans sa propre particule caractéristique. Ceci fait une différence par rapport à l'emballage d'une longueur unique d'acide nucléique dans une capsid dont la forme est déterminée. Cependant, dans le cas des virus dont les capsides ont une seule forme authentique, le mode d'assemblage présente souvent un effet indésirable. C'est la formation de certaines particules **mons-**

trueuses, aberrantes (quoique inactives) dans lesquelles la tête est plus longue que d'habitude.

Une protéine de capsid a donc une capacité intrinsèque à s'assembler en un type particulier de structure mais la taille exacte et la forme de cette structure ne sont pas nécessairement constantes. Il y a souvent d'autres **protéines (d'assemblage)**, requises pour la formation de la tête, mais qui ne font pas partie elles-mêmes de la tête ; elles peuvent limiter les options de la protéine de capsid de telle manière qu'elle ne s'assemble que selon le modèle voulu. Des génomes cellulaires peuvent également utiliser des protéines dont les fonctions sont de diriger l'assemblage d'autres protéines (voir Chapitre 26).

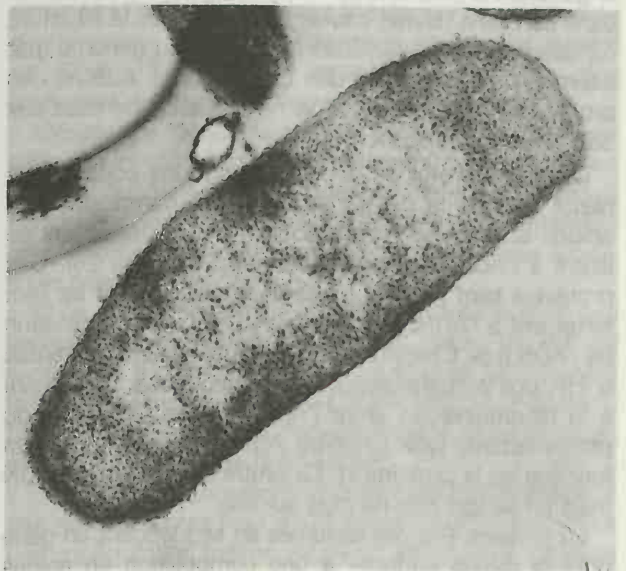
LE GÉNOME BACTÉRIEN EST UN NUCLÉOÏDE AYANT DE NOMBREUSES BOUCLES SUPERENROULÉES

Bien que les bactéries ne présentent pas de structures possédant les caractères des chromosomes eucaryotes, leurs génomes sont néanmoins organisés en corpuscules définis. On peut voir le matériel génétique sous la forme d'un amas passablement compact ou d'une série d'amas qui occupent un tiers du volume de la cellule. La **Figure 25.4** représente une section mince dans laquelle ce **nucléoïde** est bien visible.

Figure 25.4

Une section mince montre le nucléoïde bactérien sous la forme d'une masse compacte au centre de la cellule.

Photographie aimablement fournie par Jack Griffith.



Dans les bactéries qui ont partiellement répliqué leur ADN, le nucléoïde peut contenir plus d'un équivalent génomique d'ADN. Au moment de la division cellulaire, le matériel s'est disjoint en deux nucléoïdes qui se sont répartis dans les cellules filles. Le mécanisme de ségrégation implique probablement l'attachement du génome bactérien à la membrane.

Ce système représente une contrepartie de la ségrégation mitotique des chromosomes eucaryotes. Au lieu d'être attiré sur un fuseau, un locus spécifique situé sur chaque génome bactérien peut être connecté à un site sur la membrane. Les deux sites membranaires se séparent au fur et à mesure que la cellule grossit. Et, quand la division a lieu, chacun réside dans un compartiment différent, emmenant son nucléoïde avec lui (voir Chapitre 13).

Quand des cellules d'*E. coli* sont lysées, des fibres sont libérées sous la forme de boucles attachées à l'enveloppe rompue de la cellule. Comme on peut le voir sur la **Figure 25.5**, on ne trouve pas l'ADN de ces boucles sous la forme déroulée d'un double brin libre, mais il est replié dans une forme plus compacte probablement en vertu de son association à des protéines.

Plusieurs protéines se liant à l'ADN et ayant une ressemblance superficielle avec les protéines chromosomiques des eucaryotes ont été isolées chez *E. coli*. Leurs propriétés sont résumées dans le **Tableau 25.2**.

Quels critères devons-nous appliquer pour décider si une protéine se liant à l'ADN joue un rôle structural dans le nucléoïde ? D'un point de vue biochimique, elle devrait être présente en quantité suffisante pour se fixer sur tout le génome. D'un point de vue génétique, des mutations dans son gène devrait causer une rupture de la structure ou des fonctions associées à la survie du génome (par exemple, la ségrégation des cellules filles). Aucune des protéines candidates à ce rôle ne satisfait cependant aux conditions génétiques.

Deux des protéines, HU et H, sont des constituants plutôt abondants de la cellule dont la composition en acides aminés ressemble à celle des histones qui se fixent à l'ADN des eucaryotes (voir Chapitre 26). Ces protéines sont petites, hautement basiques, et se lient fortement à l'ADN. La protéine HU stimule la réplication de l'ADN (voir Chapitre 14). Sa séquence est apparentée à IHF, une protéine spécifique de séquence, nécessaire à la recombinaison entre l'ADN bactérien et l'ADN du phage lambda (voir Chapitre 28). On ne sait rien sur la fonction de la protéine H. La protéine H1 se lie à l'ADN mais on ne sait rien de plus sur elle.

La protéine P a été identifiée en séquençant un gène dont la région codante a une composition en acides

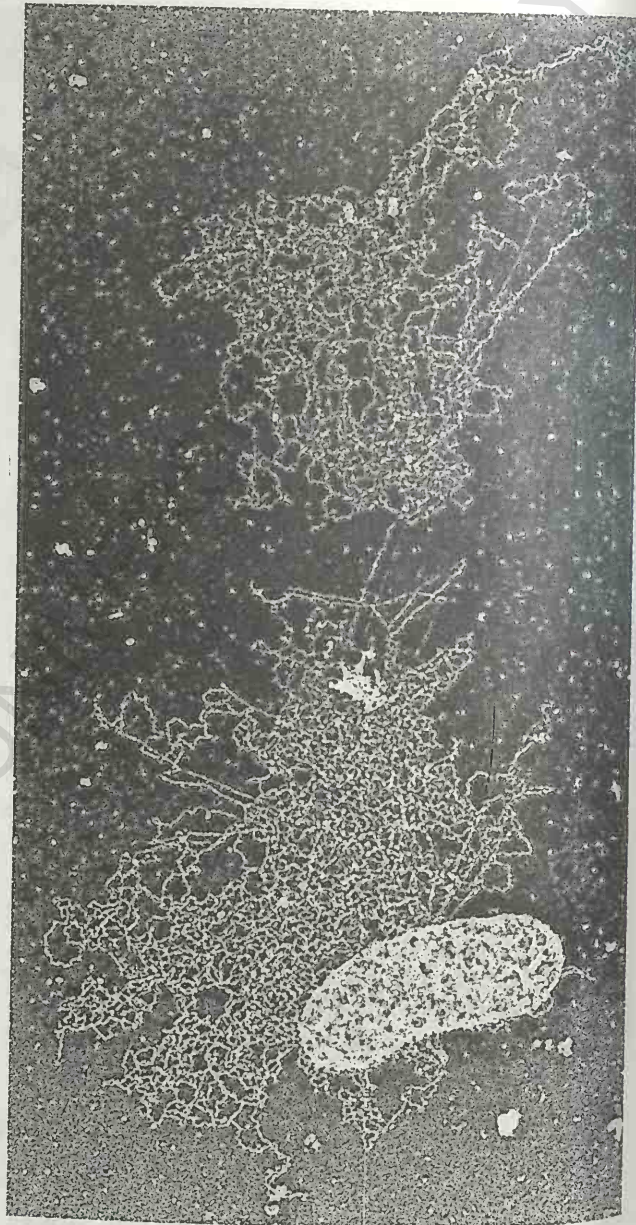


Figure 25.5

Le nucléoïde se répand à l'extérieur d'une cellule lysée d'*E. coli* sous la forme d'une fibre en boucles.

Photographie aimablement fournie par Jack Griffith.

aminés qui ressemble à celle des protamines qui se lient à l'ADN de certains spermatozoïdes. Sa séquence suggère qu'elle pourrait être une protéine se liant à l'ADN mais son abondance et sa fonction sont inconnues.

Tableau 25.2
E. Coli contient plusieurs protéines se liant à l'ADN.

Protéine	Composition	Contenu/Cellule	Équivalents eucaryotes	Locus
HU	Sous-unités α et β , chacune de 9 000 daltons	40 000 dimères	histone H2B	inconnu
H	2 sous-unités identiques de 28 000 daltons	30 000 dimères	histone H2A	inconnu
H1	Sous-unités de 15 000 daltons	10 000 copies	inconnu	inconnu
HLP1	Monomère de 17 000 daltons	20 000 copies	inconnu	<i>firA</i>
P	Sous-unités de 3 000 daltons	inconnu	protamines	inconnu

La seule protéine fermement impliquée dans une interaction avec l'appareil génétique est HLP1. Elle est codée par le locus *firA*. Des mutations peuvent abolir la résistance à la rifampicine qui est conférée par certaines mutations dans la sous-unité β de l'ARN polymérase. Les propriétés des mutants ne révèlent pas si la protéine interagit avec l'ARN polymérase ou avec l'ADN. HLP1 est petite, basique, et peut se lier à l'ADN comme les autres protéines.

Le nucléoïde peut être isolé directement sous la forme d'un complexe sédimentant très rapidement, qui consiste approximativement en 80 % d'ADN. (Les complexes analogues des eucaryotes ont environ 50 % d'ADN comme le décrit le Chapitre 26.) Il peut se déployer grâce à un traitement avec des agents actifs sur l'ARN ou les protéines. Le rôle potentiel des protéines dans la stabilisation de sa structure est clair. Le rôle que l'ARN pourrait jouer a été complètement réfractaire à l'analyse. Des tentatives pour isoler un ARN spécifique impliqué dans une activité structurale ont échoué. De même ont échoué les tentatives pour impliquer les chaînes d'ARN naissantes (c'est-à-dire les produits incomplets de la transcription en train de progresser).

L'ADN de l'élément compact isolé *in vitro* se comporte comme une structure double brin fermée, à en juger par sa réponse au bromure d'éthidium. Cette petite molécule s'intercale entre les paires de bases pour créer des tours superhélicoïdaux dans les molécules d'ADN circulaires fermées, c'est-à-dire, les molécules dans lesquelles les deux brins ont une intégrité de leurs liaisons covalentes. (Dans les molécules d'ADN circulaires « ouvertes », qui contiennent une coupure sur un brin, ou dans les molécules linéaires, l'ADN peut tourner librement en réponse à l'intercalation, réduisant ainsi la tension.)

Dans un ADN naturel fermé qui est superenroulé négativement, l'intercalation de bromure d'éthidium supprime d'abord les supertours négatifs puis introduit des supertours positifs. La quantité de bromure d'éthidium

nécessaire pour atteindre l'état de superenroulement nul est une mesure de la densité en supertours négatifs présents à l'origine.

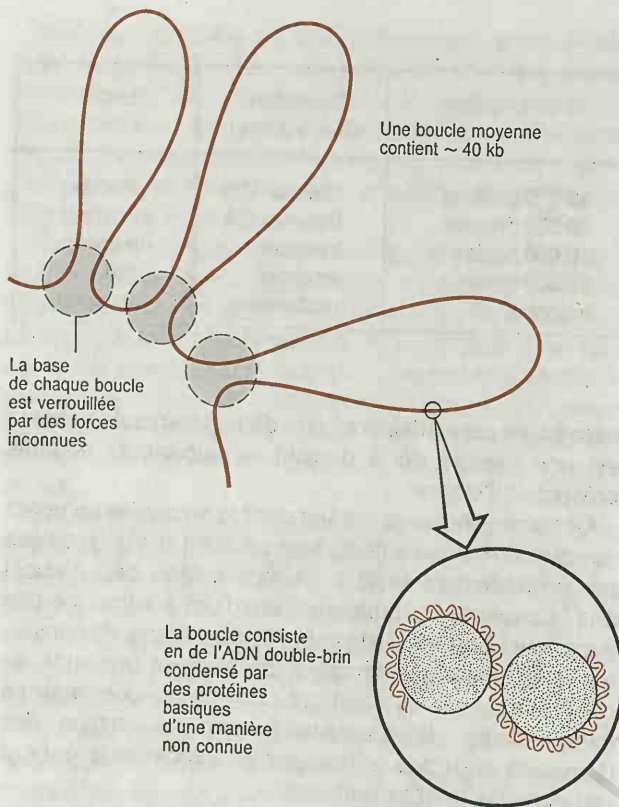
Certaines coupures ont lieu dans le nucléoïde compact au cours de son isolement ; elles peuvent être engendrées par un traitement limité à l'ADNase. Mais ceci n'abolit pas la capacité du bromure d'éthidium à introduire des supertours positifs. Cette capacité du génome de conserver la réponse au bromure d'éthidium en présence de coupures signifie qu'il doit avoir de nombreux domaines indépendants ; le superenroulement de chacun des domaines n'est pas affecté par les événements qui ont lieu dans les autres domaines.

Cette autonomie suggère que la structure du chromosome bactérien peut avoir l'organisation générale figurée sur le diagramme de la Figure 25.6. Chaque domaine consiste en une boucle d'ADN dont les extrémités sont verrouillées d'une manière (inconnue) qui ne permet pas aux événements de rotations de se propager d'un domaine à l'autre. Il y a environ 100 domaines par génome de telle sorte que chacun consiste approximativement en 40 kb (13 μ m) d'ADN organisé en une fibre plus compacte dont la structure reste à caractériser.

L'existence de domaines séparés devrait autoriser différents degrés de superenroulement dans différentes régions du génome. Ceci est un facteur pertinent à considérer pour apprécier les différentes susceptibilités de promoteurs bactériens particuliers à former des supertours (voir Chapitre 9).

Si un ADN superenroulé est libre, sa mobilité n'est pas entravée et les supertours négatifs créent un état de tension de torsion qui peut être supprimé en déroulant la double hélice, comme on le décrit dans le Chapitre 3. L'ADN peut être un état d'équilibre dynamique entre les états de tension et de déroulement.

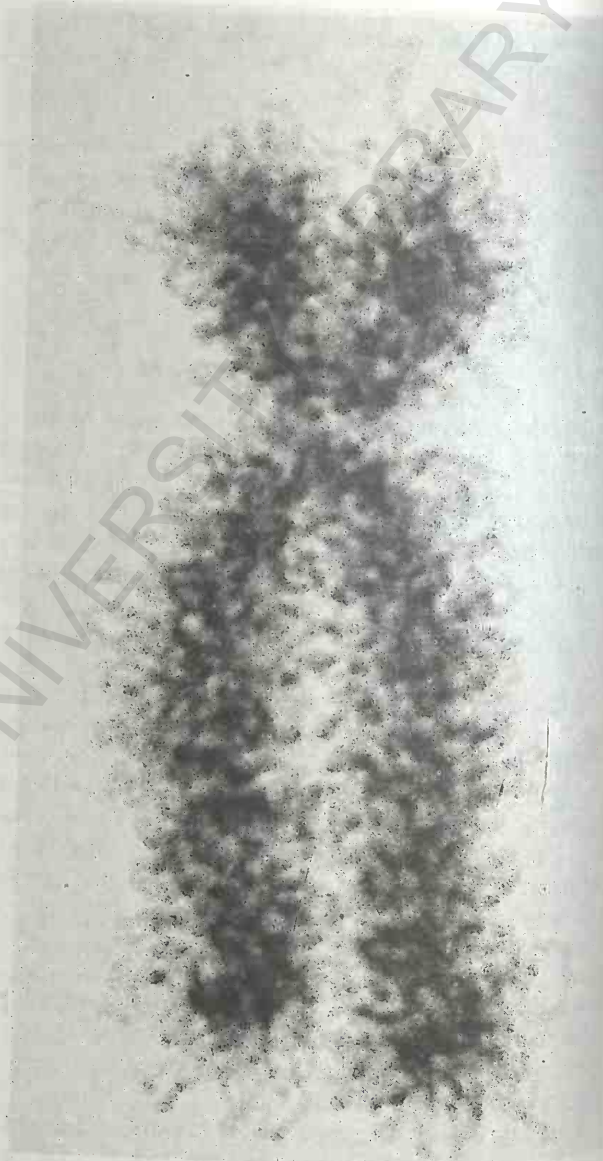
Cependant le superenroulement peut être entravé si des protéines sont liées à l'ADN pour le maintenir dans une configuration tridimensionnelle particulière. Dans ce

**Figure 25.6**

Les chromatides sœurs d'une paire mitotique consistent chacune en une fibre (d'environ 30 nm de diamètre), repliée de manière compacte pour former un chromosome.

cas les supertours sont représentés par la trajectoire que suit l'ADN dans son association fixe avec les protéines. L'énergie d'interaction entre les protéines et l'ADN superenroulé influence la stabilité de la double hélice, qui pourrait être stabilisée sous forme double brin ou sous forme simple brin de manière à ce qu'aucune tension ne soit transmise le long de la molécule.

Les supertours de l'ADN de *E. coli* sont-ils entravés *in vivo* ou bien la double hélice est-elle sujette à la tension de torsion caractéristique de l'ADN libre ? La mesure du superenroulement *in vitro* se heurte à des difficultés car des protéines qui entravent l'ADN pourraient avoir été perdues au cours de l'isolement. Cependant la situation *in vivo* peut être examinée par l'utilisation d'un agent pontant, le psoralène, qui se fixe plus rapidement à l'ADN quand celui-ci est sous tension de torsion. La réaction du psoralène avec l'ADN de *E. coli* *in vivo* correspond à une densité moyenne d'un supertour négatif toutes les

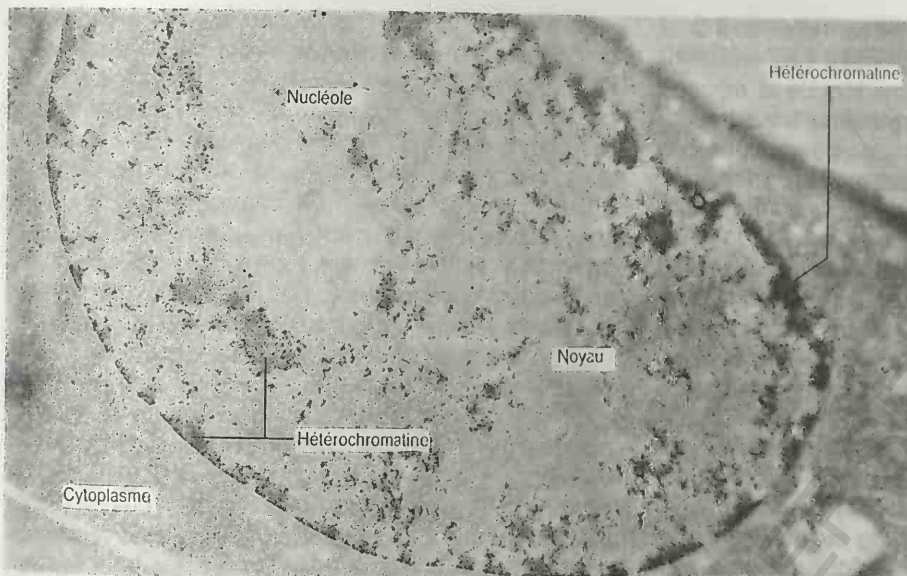
**Figure 25.7**

Les chromatides sœurs d'une paire mitotique consistent chacun en une fibre (d'environ 30 nm de diamètre), repliée de manière compacte pour former un chromosome.

Photographie aimablement fournie par E. J. Du Praw.

200 pb. Ce résultat démontre que les supertours créent bien une tension torsionnelle *in vivo*. La plupart des supertours naturels de l'ADN de *E. coli* pourraient donc ne pas être entravés.

Beaucoup de caractères importants de la structure du nucléoïde compact restent à établir. Avec quelle spécifi-

**Figure 25.8**

Une section mince dans un noyau coloré au Feulgen met en évidence l'hétérochromatine sous la forme de régions compactes regroupées près du nucléole et de la membrane nucléaire.

Photographie aimablement fournie par Edmund Puvion.

cités les domaines sont-ils construits ? Est-ce que les mêmes séquences résident toujours aux mêmes emplacements relatifs ou bien est-ce que le contenu de chacun des domaines peut changer ? Comment l'intégrité du domaine est-elle maintenue ? Quel rôle de structure jouent les protéines se liant à l'ADN ? L'analyse biochimique par elle-même pourrait être incapable de répondre pleinement à ces questions, mais s'il est possible de trouver des techniques sélectives convenables, les propriétés des mutants de structure devraient conduire à une analyse moléculaire de la construction du nucléoïde.

LE CONTRASTE ENTRE LA CHROMATINE INTERPHASIQUE ET LES CHROMOSOMES MITOTIQUES

Les chromosomes eucaryotiques individualisés n'apparaissent bien en vue que pendant une courte période au cours de la division cellulaire. Alors seulement on peut voir chacun d'entre eux sous la forme d'une unité compacte d'un rapport de condensation atteignant 10 000. La **Figure 25.7** est une photographie en microscopie électronique d'une paire de chromatides sœurs, prise en

métaphase. (Les chromatides sœurs sont des chromosomes issus de l'événement de réplication précédent, encore jointifs à ce stade de la mitose comme on l'a décrit dans le Chapitre 1.) Chacune d'elle consiste en une fibre d'un diamètre de 30 nm environ et d'apparence charbonneuse repliée dans la structure chromosomique.

Au cours de la plus grande partie du cycle de vie de la cellule eucaryote, cependant, le matériel génétique occupe une région du noyau dans laquelle on ne peut pas distinguer de chromosomes individualisés. La structure de la chromatine à l'interphase ne change pas de manière visible entre les divisions. Par exemple, on ne peut voir aucune scission pendant la période de réplication quand la quantité de chromatine double. La chromatine est fibrillaire bien que la configuration globale de la fibre dans l'espace soit difficile à discerner en détail. La fibre elle-même cependant est semblable ou identique à celle des chromosomes mitotiques.

La chromatine peut être divisée en fonction du type de matériel que l'on peut voir dans la **Figure 25.8**. Dans la plupart des régions les fibres sont bien moins condensées que dans les chromosomes mitotiques. On appelle ce matériel l'**euchromatine**. Son rapport de condensation global est probablement d'environ 1 000 à 2 000 au cours de l'interphase.

Certaines régions de la chromatine sont le siège d'une forte condensation de fibres, présentant une situation comparable à celle du chromosome lors de la mitose. On appelle ce matériel l'**hétérochromatine**. Il traverse le cycle cellulaire avec relativement peu de changement dans son degré de condensation. Souvent, les différentes régions d'hétérochromatine s'agrègent pour former un **chromocentre** dense aux colorants.

Les mêmes fibres courent continuellement entre l'euchromatine et l'hétérochromatine, ce qui implique que ces états présentent différents degrés de condensation du matériel génétique. De même, les régions euchromatiques existent dans les états différents de condensation au cours de l'interphase et au cours de la mitose. Ainsi le matériel génétique est organisé d'une manière qui permet à des états alternatifs d'être maintenus l'un à côté de l'autre dans la chromatine, et à des changements cycliques d'avoir lieu au cours de la condensation de l'euchromatine entre l'interphase et la division.

Il y a une corrélation entre la situation structurale du matériel génétique et son activité transcriptionnelle. Des chromosomes mitotiques sont transcriptionnellement inertes ; les cellules cessent virtuellement leur transcription au cours du processus de division. Il y a deux classes d'hétérochromatine, chacune contenant un type différent de séquence ; et l'ADN n'est transcrit dans aucun de ces types.

L'**hétérochromatine constitutive** consiste en des régions particulières qui ne sont *jamais* exprimées. Elle inclut des séquences d'ADN satellites. L'**hétérochromatine facultative** prend la forme de chromosomes entiers qui sont inactifs dans un lignage cellulaire bien qu'ils puissent être exprimés dans d'autres circonstances. L'exemple « *par excellence* » est le chromosome X des mammifères dont une copie (choisie au hasard) est entièrement inactive chez une femelle. (Ceci compense la présence de deux chromosomes X, alors qu'il n'y a qu'un seul X présent chez les mâles.) Le chromosome X inactif est perpétué dans un état hétérochromatique ; le chromosome X actif fait partie de l'euchromatine. Il est possible ici de voir une corrélation entre l'activité transcriptionnelle et l'organisation structurale quand des *séquences d'ADN identiques* existent dans les deux états.

Une très forte condensation du matériel génétique est ainsi associée à son inactivité (et pourrait en être responsable). Notez, cependant, que la réciprocité n'est pas vraie. Des gènes actifs sont en effet contenus dans l'euchromatine ; mais, seule, une petite minorité des séquences de l'euchromatine est transcrite à chaque moment. Ainsi le fait d'être situé dans l'euchromatine est *nécessaire* à l'expression du gène, mais n'est pas *suffi-*

sant. Nous pouvons nous demander si les différences grossières que nous voyons entre l'euchromatine et l'hétérochromatine sont reproduites à un moindre degré par des changements dans la structure de l'euchromatine, qui donnent aux régions transcrites une structure moins condensée que celle des régions non transcrites.

Quand on les isole *in vitro*, la chromatine interphasique et les chromosomes mitotiques possèdent de larges boucles de la fibre élémentaire dont ils sont constitués, apparemment verrouillées à la base pour former des domaines indépendants, analogues à ceux que l'on trouve dans le nucléoïde bactérien. Leur structure est discutée dans le Chapitre 26.

En raison de l'état diffus de la chromatine, nous ne pouvons pas déterminer directement la spécificité de son organisation. Mais nous pouvons nous demander quel est le niveau d'organisation de la structure du chromosome. Est-ce que des séquences particulières résident à des sites particuliers, ou bien est-ce que le repliement de la fibre à l'intérieur de la structure globale est un événement plus aléatoire ?

Il y a quelques années, le développement de la technique de **coloration des bandes G (G-banding)** a montré que chaque chromosome a une ultrastructure différente et reproductible. Quand ils sont soumis à certains traitements puis colorés avec le colorant chimique Giemsa, les chromosomes produisent une série de bandes. Un exemple de l'assortiment chromosomique humain est présenté dans la **Figure 25.9**.

Avant le développement de cette technique, les chromosomes pouvaient seulement être distingués par leur taille globale, et l'emplacement relatif de leur centromère (voir plus loin). Maintenant, chaque chromosome peut être identifié par son mode de coloration en bandes caractéristique. Ce mode caractéristique est suffisamment reproductible pour permettre l'identification de translocations d'un chromosome à l'autre par comparaison avec l'assortiment diploïde normal.

La technique de coloration des bandes est d'une utilité pratique considérable, mais le mécanisme de la coloration des bandes reste un mystère. Tout ce qui est certain, c'est que le colorant marque des chromosomes non traités de manière plus ou moins uniforme. Ainsi la production de bandes dépend d'une variété de traitements qui modifient la réponse du chromosome (probablement en extrayant des régions non marquées le constituant qui fixe le colorant) mais la variété des traitements est si grande qu'aucune cause commune n'a encore été discernée. Ces résultats impliquent qu'il existe une structure définie sur une longue distance, mais dont la base est inconnue.

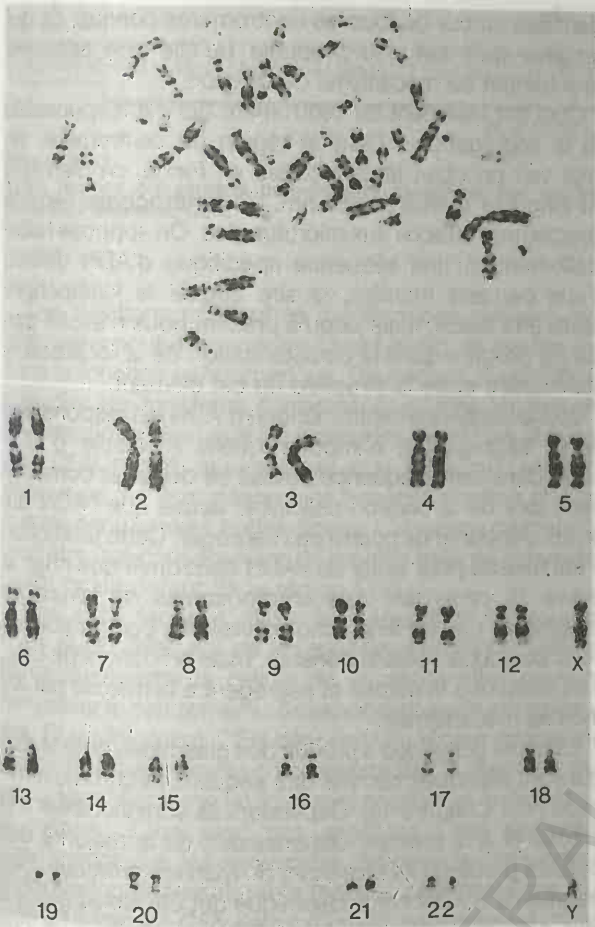


Figure 25.9
La coloration des bandes G crée une série caractéristique de bandes transversales sur chaque membre de l'assortiment chromosomique.
Cet exemple d'un caryotype humain a été fourni par N. Davidson.

Chaque chromosome contient un double brin d'ADN unique, très long. Ceci explique pourquoi la réplication chromosomique est semi-conservative, comme celle de la molécule individuelle d'ADN (voir Figure 2.14) ; ceci aurait peu de chance d'être le cas si un chromosome comportait plusieurs molécules indépendantes d'ADN. Le double brin unique d'ADN est replié pour former la fibre de 30 nm qui s'étend en ligne continue d'un bout à l'autre du chromosome. Ainsi, pour rendre compte de la structure de la chromatine à l'interphase et de celle du chromosome mitotique, nous devons expliquer la condensation d'une seule molécule extrêmement longue d'ADN, sous une forme dans laquelle elle peut être transcrite et répliquée, et peut devenir cycliquement plus ou moins compacte.

LES CHROMOSOMES EUCARYOTES COMME MOYEN DE SÉGRÉGATION

Au cours de la mitose, les chromatides sœurs se déplacent vers les pôles opposés de la cellule (comme l'illustre la Figure 1.6). Leur mouvement dépend de l'attachement du chromosome aux microtubules dont l'autre extrémité est connectée aux pôles. [Les microtubules sont formés d'un système filamenteux cellulaire réorganisé à la mitose de telle manière qu'une extrémité de chacun des microtubules est jointe aux centrioles (ou à un point voisin) qui résident à chaque pôle de la cellule.]

Le site au niveau duquel un chromosome est attaché aux microtubules, est identifié par une région de constriction appelée **centromère**. Cette constriction est nettement visible sur la photographie de la Figure 25.7, qui montre les chromatides sœurs au stade de la métaphase de la mitose alors qu'elles sont encore attachées l'une à l'autre dans la région centromérique.

Historiquement, le terme centromère a été utilisé à la fois dans le sens fonctionnel et structural pour décrire le déterminant du chromosome responsable de son mouvement. Le centromère est attiré vers les pôles au cours de la mitose, et le chromosome qui lui est attaché est en quelque sorte entraîné. De ce point de vue, le chromosome est un moyen pour attacher un grand nombre de gènes à l'appareil de division.

Le centromère est essentiel pour la ségrégation, comme le montre le comportement des chromosomes qui ont été cassés. Une cassure unique engendre un fragment qui garde le centromère et un autre **fragment acentrique** qui en est dépourvu. Le fragment acentrique ne s'attache pas au fuseau mitotique, et en conséquence, il peut manquer d'être inclus dans l'un des deux noyaux engendrés.

(Quand le mouvement chromosomique repose sur des centromères individualisés, il ne peut y avoir qu'un *seul* centromère par chromosome. Si des translocations produisent des chromosomes qui ont plusieurs centromères, des structures aberrantes se forment à la mitose puisque les deux centromères situés sur la même chromatide sœur peuvent être attirés vers des pôles différents, cassant le chromosome. Cependant, dans certaines espèces, les centromères sont « diffus », ce qui crée une situation différente. Les centromères individualisés sont les seuls qui aient été analysés au niveau moléculaire.)

Les régions qui bordent le centromère sont souvent riches en séquences satellites et peuvent contenir une quantité considérable d'hétérochromatine constitutive. Comme le chromosome entier est condensé, l'hétérochromatine centromérique n'est pas immédiatement visi-

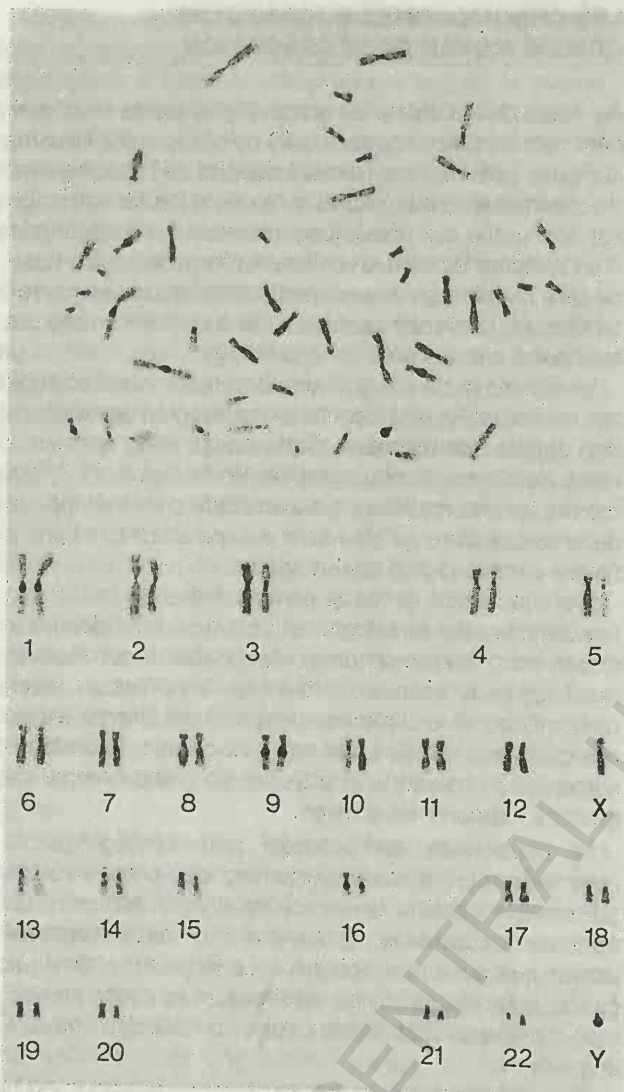


Figure 25.10

La coloration des bandes C produit une coloration intense des centromères de tous les chromosomes.

Photographie aimablement fournie par N. Davidson.

ble dans les chromosomes mitotiques, mais peut être visualisée par la technique du « C-banding » (coloration des bandes C). Dans l'exemple de la Figure 25.10, tous les centromères apparaissent comme des régions prenant fortement le colorant. Bien qu'elle soit très commune, l'hétérochromatine constitutive ne peut pas être

identifiée autour de tous les centromères connus, ce qui suggère qu'il est peu probable qu'elle soit associée directement au mécanisme de division.

Quel est l'élément du centromère qui est responsable de la ségrégation ? Dans la région du centromère, on peut voir un objet fibreux coloré en foncé, de diamètre ou longueur d'environ 400 nm. Ce **kinétochore** semble directement attaché aux microtubules. On suppose habituellement qu'une séquence spécifique d'ADN définit, d'une certaine manière, le site auquel le kinétochore devra être établi, mais jusqu'à présent nous n'avons pas fait de progrès dans la caractérisation de la localisation moléculaire et de la structure de cet élément.

Si une séquence centromérique d'ADN est responsable de la ségrégation, n'importe quelle molécule d'ADN possédant cette séquence devrait se déplacer correctement lors de la division cellulaire, tandis que l'ADN qui en est dépourvu ne pourra pas ségréger. Cette prédiction a été utilisée pour isoler de l'ADN centromérique chez la levure *S. cerevisiae*. Les chromosomes de levure ne présentent pas de kinétochores visibles, comparables à ceux des eucaryotes supérieurs, mais en dehors de cela, il se divisent à la mitose et ségrègent à la méiose par les mêmes mécanismes.

Le génie génétique a produit des plasmides de la levure qui sont répliqués comme des séquences chromosomiques (voir Chapitre 13). Cependant ils sont instables à la mitose et à la méiose, disparaissant de la majorité des cellules parce qu'ils ségrègent de manière erratique. Des fragments d'ADN chromosomique ont été isolés grâce à leur capacité à conférer la stabilité mitotique à ces plasmides. Leur origine chromosomique peut être identifiée quand ils contiennent des marqueurs génétiques déjà cartographiés près d'un centromère.

En réduisant la taille des fragments qui sont incorporés dans le plasmide, la longueur minimum nécessaire à la fonction de centromère mitotique peut être identifiée. Plusieurs fragments *CEN* ont été caractérisés. Ils ont en commun trois types de séquences qui sont résumées dans la Figure 25.11.

L'élément I est une courte séquence représentée avec quelques variations dans tous les centromères. L'élément II est une séquence riche en A-T (plus de 90 %), de 78 à 86 paires de bases, retrouvée dans tous les centromères ; sa fonction pourrait dépendre de sa longueur, plutôt que de sa séquence exacte. Sa constitution rappelle certains ADN satellites des eucaryotes supérieurs. L'élément III est une région courte, bien conservée dans tous les centromères connus. Aucune des séquences centromériques n'a de cadre ouvert de lecture.

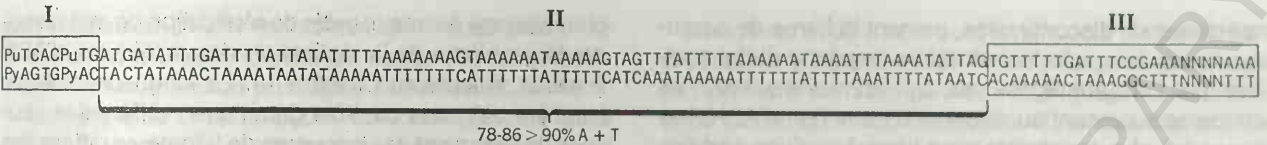


Figure 25.11
Trois régions conservées peuvent être identifiées par les homologues de séquence entre les différents éléments *CEN*.

Des délétions et d'autres modifications peuvent être effectuées pour examiner les caractéristiques impliquées dans la fonction centromérique. Cependant, il est difficile d'étudier les fonctions centromériques directement sur les plasmides. Bien que les plasmides ayant des séquences *CEN* soient plus stables que ceux qui en sont dépourvus, ils sont néanmoins beaucoup moins stables que les centromères authentiques. L'absence de stabilité peut être une conséquence de la petite taille du plasmide, plutôt qu'un défaut de la région *CEN*.

Une autre façon d'utiliser les séquences centromériques disponibles est de les modifier *in vitro* puis de les réintroduire dans la cellule de levure où elles peuvent remplacer le centromère correspondant sur le chromosome. Si un fragment *CEN* dépourvu de région conservée remplace le centromère chromosomique, le chromosome se comporte comme un fragment acentrique au cours de la mitose : il ne peut pas ségréger. Ceci démontre directement que les séquences conservées comportent l'information nécessaire à la fonction centromérique.

Si un fragment *CEN* dérivé d'un chromosome est utilisé pour remplacer le centromère d'un autre chromosome aucun changement n'est observé. Ce résultat suggère que les centromères sont interchangeables. Ils sont utilisés seulement pour attacher les chromosomes au

fuseau, et ne jouent aucun rôle pour distinguer un chromosome d'un autre.

Une autre caractéristique, qui est essentielle pour tous les chromosomes, est le **télomère**. D'une certaine manière que nous ne comprenons pas encore, celui-ci scelle l'extrémité. On sait depuis longtemps que l'extrémité doit être une structure particulière, car les extrémités chromosomiques produites par cassure sont collantes et ont tendance à réagir avec d'autres chromosomes, tandis que les extrémités naturelles sont stables.

Nous pouvons appliquer deux critères pour identifier une séquence télomérique. Premièrement, elle doit résider à l'extrémité d'un chromosome (ou au moins à l'extrémité d'une molécule d'ADN linéaire authentique). Deuxièmement, elle doit conférer la stabilité à une molécule linéaire.

Plusieurs séquences télomériques ont été obtenues à partir de molécules d'ADN linéaires présentes dans le génome d'eucaryotes inférieurs. Chaque télomère consiste en une courte séquence répétée en tandem. Le **Tableau 25.3** dresse la liste des unités répétées qui ont été identifiées aux extrémités des molécules d'ADN linéaires. On peut toutes les écrire sous la forme $C_n(A/T)_m$ où n est supérieur à 1 et m vaut de 1 à 4.

On trouve dans la région télomérique un arrangement

Tableau 25.3
Les télomères ont en commun une courte répétition en tandem.

Type d'organisme	Espèce	Source d'ADN	Unité répétée (5'-3')
Holotriches (ciliés) Hypotriches (ciliés) Flagellés Moississures de vase Levure	<i>Tetrahymena</i> , <i>Paramecium</i> <i>Stylonchia</i> , <i>Oxytricha</i> <i>Trypanosoma</i> , <i>Leishmania</i> <i>Physarum</i> , <i>Dictyostelium</i> <i>Saccharomyces</i>	macronoyau macronoyau minichromosome ADNr chromosome	CCCCAA CCCCAAAA CCCTA CCCTA _n C ₂₋₃ A(CA) ₁₋₃

L'unité de répétition donne la séquence d'une chaîne allant du télomère vers le centre.

spécifique de discontinuités, prenant la forme de cassures simple brin, que leur structure empêche d'être scellées par l'enzyme ligase qui agit normalement sur les coupures survenant au niveau d'un seul brin d'ADN. Les bases les plus terminales sont bloquées d'une certaine façon — elles pourraient être organisées en épingle à cheveux — de telle manière qu'elles ne sont pas reconnues par les nucléases.

Le problème de la réalisation d'un test fonctionnel a de nouveau été porté au niveau moléculaire grâce à l'utilisation de la levure. Tous les plasmides qui survivent chez la levure sont des molécules d'ADN circulaires. Les plasmides linéaires sont instables (car ils sont dégradés). Est-ce qu'une séquence authentique d'ADN télomérique pourrait conférer la stabilité à un plasmide linéaire ?

Une région de l'extrémité d'une molécule naturelle connue d'ADN linéaire — l'ADNr extrachromosomique de *Tetrahymena* — est capable de rendre stable un plasmide de levure sous sa forme linéaire. Les coupures simple brin dans la séquence télomérique sont perpétuées aux mêmes sites chez la levure, ce qui constitue une remarquable conservation entre les espèces.

La séquence de *Tetrahymena* peut être remplacée par des fragments de l'ADN de levure qui ont des séquences apparentées. Il est probable qu'elles se révéleront être des télomères naturels des chromosomes de levure. Le même type d'analyse que celui effectué avec l'ADN centromérique, utilisant des fragments de taille réduite et l'induction de mutations, devrait révéler la manière dont ils fonctionnent.

Certaines indications sur la façon dont un télomère fonctionne sont données par certaines propriétés inhabituelles des extrémités des molécules d'ADN linéaires. Dans une population de trypanosomes, les extrémités sont de longueurs variables. Quand on suit la croissance d'un clone individuel, le télomère croît de 7 à 10 paires de bases (une à deux répétitions) par génération. Encore plus révélateur est le devenir des télomères de ciliés introduits chez la levure. Après la réplication dans la levure, des répétitions télomériques de levure sont ajoutées aux extrémités des répétitions de *Tetrahymena*.

Comment sont synthétisées les répétitions télomériques ? Les extrémités de *Tetrahymena* contiennent une activité enzymatique qui utilise le 3'-OH du brin télomérique G + T comme amorce pour la synthèse de répétitions en tandem TTGGGG. Seuls le dGTP et le dTTP sont nécessaires pour l'activité. Nous ne savons pas comment le brin complémentaire du télomère est assemblé.

L'addition de répétitions télomériques à l'extrémité du

chromosome à chaque cycle de réplication devrait résoudre le problème de la réplication des molécules d'ADN linéaires, discuté au Chapitre 14. La longueur totale du télomère est sous contrôle génétique ; différentes souches de levure ont des longueurs de télomères différentes mais caractéristiques. Il doit donc exister un mécanisme pour empêcher les extrémités de devenir trop longues, qui consiste peut-être en l'enlèvement de certaines des répétitions. Un gène essentiel de la levure empêche les télomères de s'étendre ; quand il est muté, ils grandissent sans arrêt.

Si les télomères sont continuellement allongés (ou raccourcis), leur séquence exacte pourrait être sans intérêt. La seule chose requise est que l'extrémité présente soit reconnue comme substrat convenable pour l'addition. Ceci explique pourquoi le télomère de cilié fonctionne chez la levure. Cependant, nous ne comprenons pas encore comment la séquence télomérique est en rapport avec la résistance des extrémités chromosomiques aux dommages.

Les caractéristiques minimum nécessaires à l'existence d'un chromosome sont la possession d'un télomère convenable pour assurer la survie, un centromère pour permettre la ségrégation, et une origine où la réplication est initiée (voir Chapitre 13). Tous ces éléments ont été rassemblés pour construire un chromosome synthétique de levure. Il s'est avéré que le chromosome synthétique n'est stable que s'il est plus long que 20-50 kb. Nous ne connaissons pas la base de ce phénomène, mais la possibilité de construire un chromosome synthétique offre le moyen d'examiner la nature du système de ségrégation dans un environnement contrôlé*.

CERTAINS GÈNES SONT EXTRACHROMOSOMIQUES

Il existe deux situations dans lesquelles les unités de transcription de l'ARNr ne résident pas dans le chromosome, mais sont présentes en tant que molécules extrachromosomiques indépendantes.

On trouve une telle situation de manière stable chez certains eucaryotes inférieurs, chez lesquels les gènes responsables de la production d'ARNr sont portés par

Figure 25.12

Les gènes des ARNr de *D. discoideum* résident dans un palindromes extrachromosomique.



* N. d. T. : La structure des télomères de nombreux eucaryotes, et notamment de l'homme, a récemment été caractérisée. Tous les télomères eucaryotiques semblent posséder le même type de répétition que ce qui est décrit ici pour *Tetrahymena*. Il semble qu'existe, au niveau de ces télomères, un appariement spécial n'obéissant pas aux règles de Watson et Crick et impliquant des paires GG.

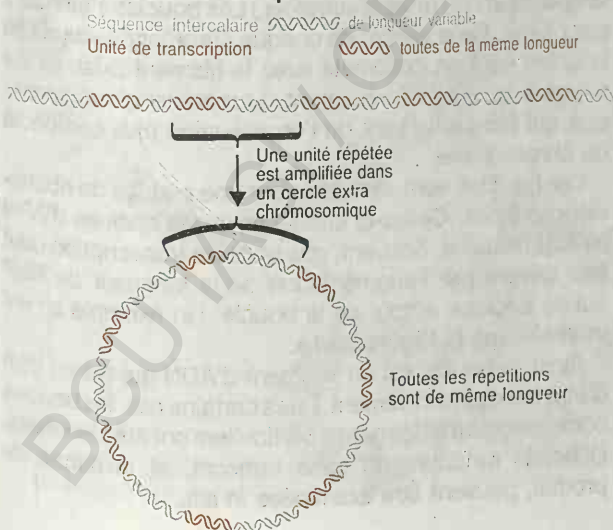
des molécules d'ADN extrachromosomiques présentes en de nombreux exemplaires par noyau. (De telles molécules sont une source privilégiée de télomères.)

L'exemple de *D. discoideum* est montré dans la Figure 25.12. La molécule extrachromosomique est un grand dimère palindromique contenant deux unités de transcription en orientation opposée, séparées par environ 20 000 paires de bases. Chaque dimère extrachromosomique contient également deux gènes 5S, bien que ceux-ci soient transcrits indépendamment de l'unité majeure de transcription. Nous savons très peu de choses sur les arrangements permettant l'héritage de cette molécule.

Un état transitoire a lieu au cours du développement de certains animaux. Au cours de l'ovogenèse, dans de nombreuses espèces, un grand nombre de gènes *additionnels* d'ARNr sont soudainement produits. Le processus a été caractérisé de la façon la plus complète chez *X. laevis*. Des molécules circulaires extrachromosomiques d'ADN contenant chacune de nombreuses unités répétées sont produites par **amplification** d'unités de répétition chromosomiques.

Contrairement aux gènes chromosomiques, dans lesquels des unités répétées peuvent avoir des séquences intercalaires de tailles différentes, sur les cercles extrachromosomiques, les unités adjacentes ont toutes des séquences intercalaires de *même* longueur. Ceci suggère que chaque cercle amplifié est produit en fabriquant de multiples copies consécutives d'une seule unité chromosomique, comme l'illustre la Figure 25.13.

Figure 25.13
Chaque ADN extrachromosomique dans l'oocyte de *X. laevis* est un cercle engendré par l'amplification en série d'une seule unité répétée d'ADNr chromosomique.



Les gènes additionnels sont requis pour permettre une production suffisante d'ARNr dans l'oocyte et ne sont pas présents dans les cellules somatiques, qui fonctionnent seulement avec les gènes chromosomiques. (Le mécanisme de cette amplification est une forme de réplication que l'on appelle le cercle roulant, illustré dans la Figure 13.10.)

L'ÉTAT DÉROULÉ DES CHROMOSOMES EN ÉCOUVILLON

Il serait extrêmement utile de visualiser l'expression des gènes à l'état naturel, pour voir quels changements structuraux sont associés à la transcription. Mais la nature du matériel restreint une telle analyse à certaines circonstances inhabituelles.

La compression de l'ADN dans la chromatine, couplée à la difficulté d'identifier des gènes particuliers en son sein, rendent impossible la visualisation de la transcription de gènes actifs considérés isolément. (Cependant, ces gènes présentent certaines propriétés biochimiques distinctes qui peuvent être analysées *in vitro*, comme le décrit le Chapitre 27.)

Les chromosomes mitotiques sont inertes du point de vue de l'expression génique, et de toute façon, ils sont tellement compacts que cela empêche l'identification de loci distincts. Les régions distinctes qui sont rendues discernables par la technique de coloration des bandes G, comme dans l'exemple de la Figure 25.9, contiennent chacune de l'ordre de 10^7 paires de bases d'ADN, ce qui pourrait inclure plusieurs centaines de gènes.

On peut voir une différenciation latérale de la structure de nombreux chromosomes lorsqu'ils apparaissent pour la première fois à la méiose. A ce stade, les chromosomes ressemblent à une série de perles enfilées. Les perles sont des granules fortement colorés, justement connus sous le nom de **chromomères**. Cependant, habituellement, il n'y a pas d'expression génétique à la méiose, et il n'est pas pratique d'utiliser ce matériel pour identifier des gènes isolés.

L'expression génétique peut être visualisée directement dans deux situations, dans lesquelles on trouve les chromosomes dans une forme hautement déroulée qui autorise la distinction de loci isolés (ou de groupes de loci). Les **chromosomes en écouvillon** ont été bien caractérisés chez certains amphibiens ; les **chromosomes polytènes** ont été bien caractérisés chez certains insectes.

Les chromosomes en écouvillon sont formés au cours d'une méiose inhabituellement longue, qui peut même durer plusieurs mois ! Au cours de cette période, les chromosomes sont maintenus dans une forme étirée

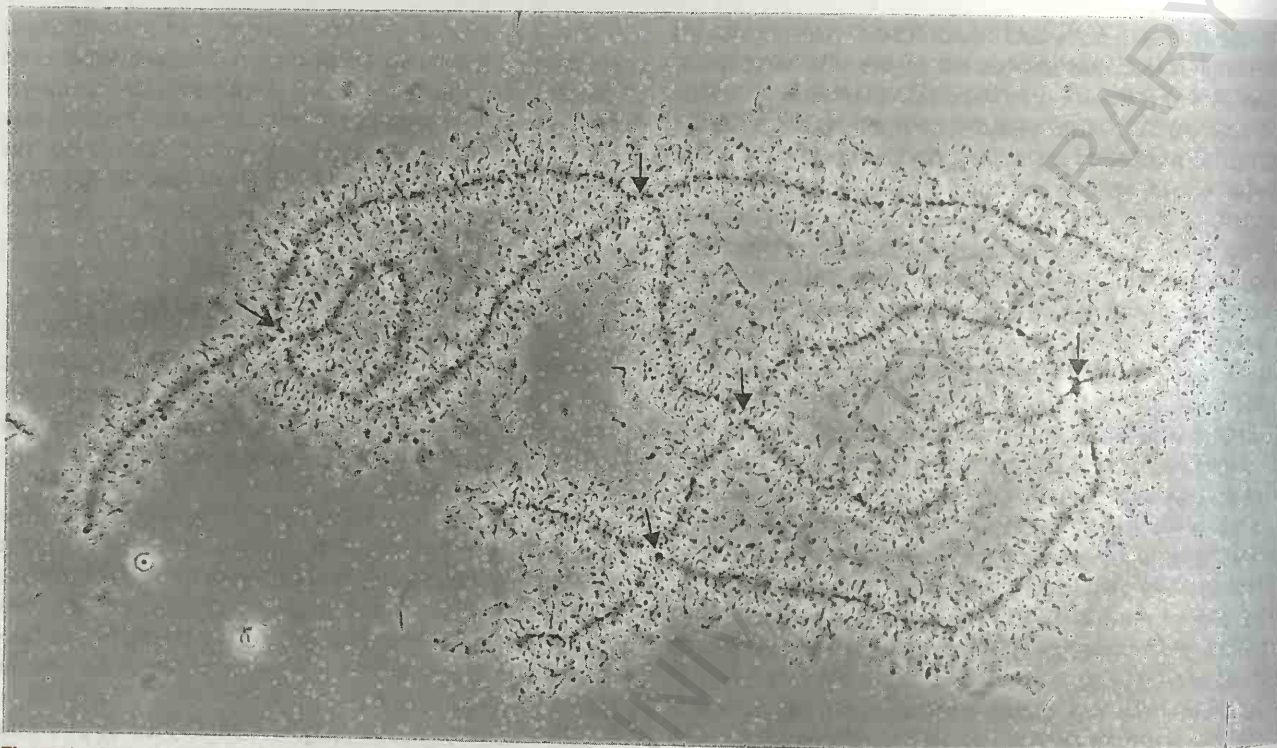


Figure 25.14

Un chromosome en écouvillon est un bivalent méiotique dans lequel les deux paires de chromatides sœurs sont maintenues ensemble au niveau des chiasmas (indiqués par des flèches). Photographie aimablement fournie par Joe Gall.

dans laquelle ils peuvent être visualisés au microscope optique. Plus tard dans la méiose, les chromosomes reviennent à leur forme compacte habituelle. Ainsi, l'état étiré offre une version dépliée de la condition normale du chromosome.

Les chromosomes en écouvillon sont des bivalents méiotiques, chacun consistant en deux paires de chromatides sœurs. La **Figure 25.14** montre un exemple où les paires de chromatides sœurs sont presque entièrement séparées de manière à ce qu'elles tiennent ensemble seulement par des chiasmas (les sites de crossing-over). Chaque paire de chromatides sœurs forme une série de chromomères ellipsoïdaux d'environ 1 à 2 microns de diamètre, qui sont connectés par un filament très fin. Ce filament contient les molécules sœurs d'ADN double brin, et court continuellement le long du chromosome, à travers les chromomères.

La longueur de chacun des chromosomes en écouvillon chez le triton *Notophthalmus viridescens* va de 400 à 800 microns, ce qui peut être comparé avec l'ordre de grandeur de 15 à 20 microns, que l'on voit plus tard dans la méiose. Ainsi les chromosomes en écouvillon sont environ 30 fois moins fortement condensés. La longueur

totale de l'assortiment entier de chromosomes en écouvillon peut être de 5 à 6 mm, organisé en 5 000 chromomères environ.

Les chromosomes en écouvillon tirent leur nom de leurs boucles latérales qui jaillissent des chromomères à certains endroits. (Celles-ci ressemblent aux plumeaux à lampes que l'on utilisait autrefois.) Les boucles s'étendent par paires, l'une venant de chaque chromatide sœur. Les boucles sont en continuité avec le filament axial, ce qui suggère qu'elles représentent du matériel chromosomique qui fait saillie hors de l'organisation plus compacte du chromomère.

Les boucles sont entourées par une matrice de ribonucléoprotéines. Celles-ci contiennent des chaînes d'ARN à l'état naissant. Souvent, une unité de transcription peut être définie par l'augmentation de la longueur de RNP qui se déplace autour de la boucle. Un exemple en est montré dans la **Figure 25.15**.

Ainsi, la boucle est un segment d'ADN qui est en train d'être transcrit activement. Dans certains cas, les boucles correspondant à des gènes particuliers ont été identifiées. Donc, la structure du gène transcrit, et la nature du produit, peuvent être examinées *in situ*.

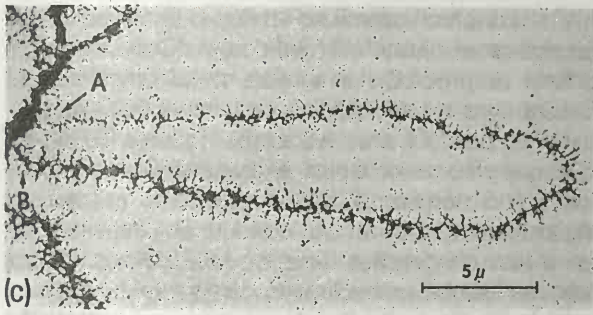


Figure 25.15

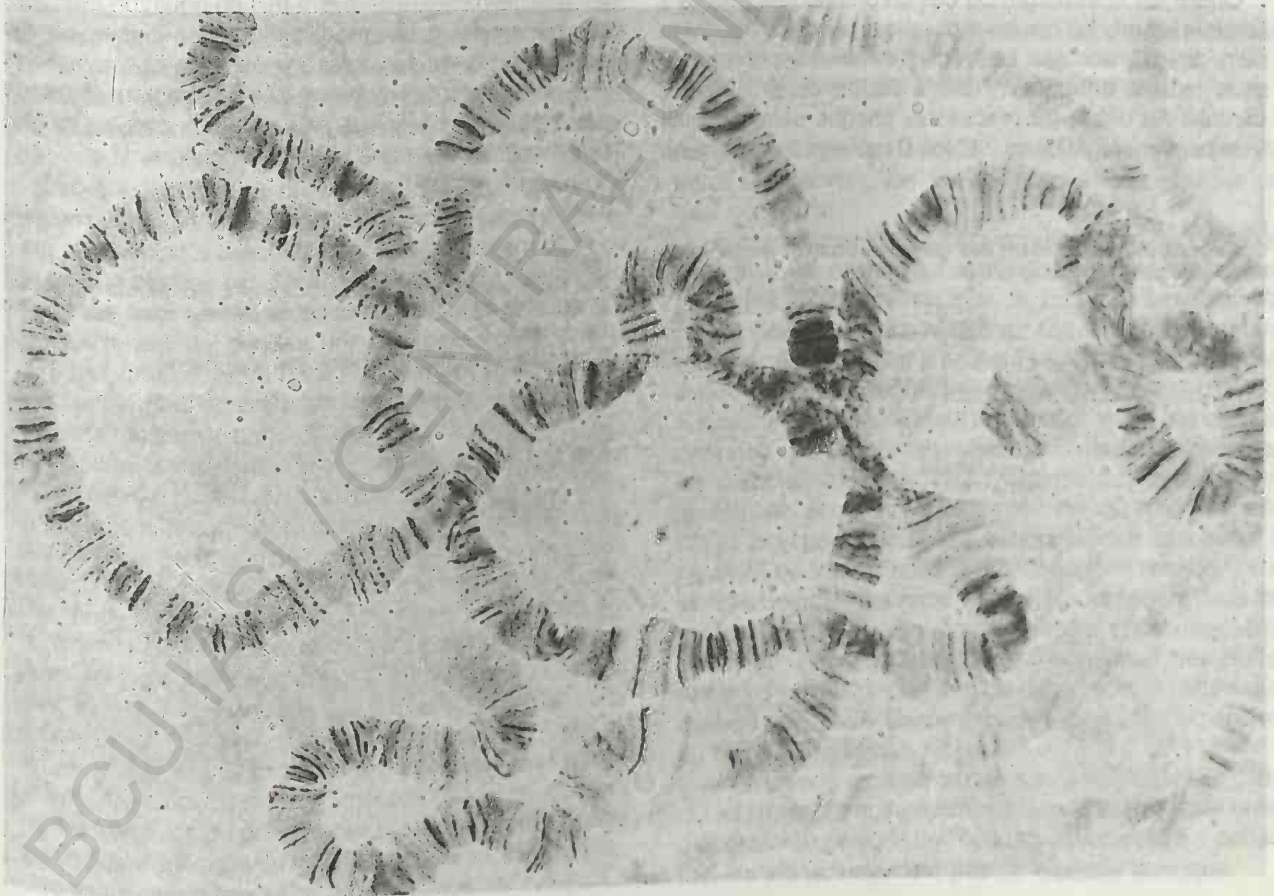
Une boucle de chromosome en écouvillon est entourée par une matrice de ribonucléoprotéine.

Photographie aimablement fournie par Oscar Miller.

Figure 25.16

Les chromosomes polytènes de *D. melanogaster* forment une série de bandes qui alternent avec des interbandes.

Photographie aimablement fournie par Jose Bonner.



LA POLYTÉNIE CRÉE DES CHROMOSOMES GÉANTS

Les noyaux interphasiques de certains tissus de larves de Diptères (mouches) contiennent des chromosomes qui sont extrêmement agrandis par rapport à leur aspect habituel. Ils ont à la fois un diamètre augmenté et une plus grande longueur. La **Figure 25.16** montre un exemple d'un assortiment chromosomique de la glande salivaire de *D. melanogaster*. Ces chromosomes sont appelés **chromosomes polytènes**.

Chaque chromosome consiste en une série de **bandes** visibles (décrites de manière plus appropriée, mais rarement comme des chromomères). Les bandes sont de taille variable allant d'une largeur d'environ 0,5 microns pour la plus grande, à environ 0,05 micron pour la plus petite. (Les plus petites peuvent être distinguées seule-

ment au microscope électronique.) Les bandes contiennent la plus grande partie de la masse d'ADN et se colorent de manière intense par les réactifs appropriés. Les régions qui les séparent sont colorées plus légèrement et sont appelées **interbandes**. Il y a environ 5 000 bandes dans l'assortiment chromosomique de *D. melanogaster*.

Les centromères des quatre chromosomes de *D. melanogaster* s'agrègent pour former un chromocentre qui consiste surtout en de l'hétérochromatine (chez le mâle, il contient le chromosome Y en entier). Déduction faite de ceci, environ 75 % de l'ADN haploïde sont organisés sous la forme d'alternance de bandes et d'interbandes.

La longueur de l'assortiment chromosomique est d'environ 2 000 μm ; 75 % de l'ADN représentent $1,3 \times 10^8$ paires de bases, qui seraient déroulées sur à peu près 40 000 μm de telle sorte que le rapport de condensation est d'environ 20. Ceci démontre clairement l'existence d'un déroulement du matériel génétique par rapport aux états habituels de la chromatine à l'interphase ou dans les chromosomes mitotiques.

Quelle est la structure de ces chromosomes géants ? Chacun est produit par les réplifications successives d'une paire diploïde accolée. Les répliques ne se séparent pas, mais restent attachées l'une à l'autre dans leur état déroulé. Au début du processus, chaque paire accolée à un contenu en ADN de 2 C (où C représente le contenu

en ADN d'un chromosome individuel). Puis celui-ci double successivement jusqu'à 9 fois, ce qui donne à la fin, un contenu de 1 024 C. Le nombre de doublings est différent dans les divers tissus de la larve de *D. melanogaster*, et, bien sûr, chez les autres Diptères.

Chaque chromosome peut être visualisé sous la forme d'un grand nombre de fibres parallèles qui courent longitudinalement, fortement condensées dans les bandes, moins condensées dans les interbandes. Chaque fibre représente probablement un seul chromosome haploïde (C). Ceci est à l'origine du nom « polyténie ». Le degré de polyténie est le nombre de chromosomes haploïdes contenus dans le chromosome géant.

Le mode de coloration en bandes est caractéristique de chaque souche de *Drosophile*. Ce nombre constant et l'agencement linéaire des bandes ont été remarqués pour la première fois dans les années 1930, et on a réalisé alors que cet agencement formait une **carte cytologique** des chromosomes. Des réarrangements tels que des délétions, inversions ou duplications résultent en des modifications de l'ordre des bandes. L'agencement linéaire des bandes peut être mis en équation avec celui des gènes. Ainsi, des modifications génétiques, comme celles que l'on observe sur une carte de liaison, peuvent être corrélées à des modifications de structure de la carte cytologique. En définitive, une mutation particulière peut être localisée dans une bande particulière.

Figure 25.18

Certaines bandes contenant des gènes particuliers peuvent être identifiées par hybridation *in situ*.

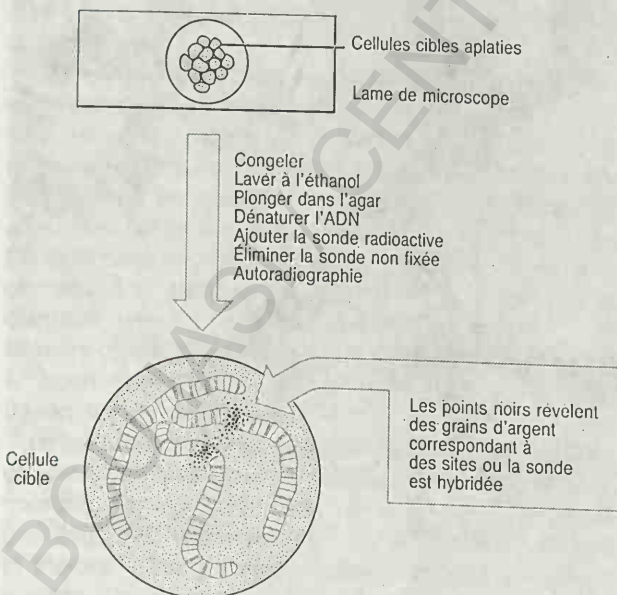
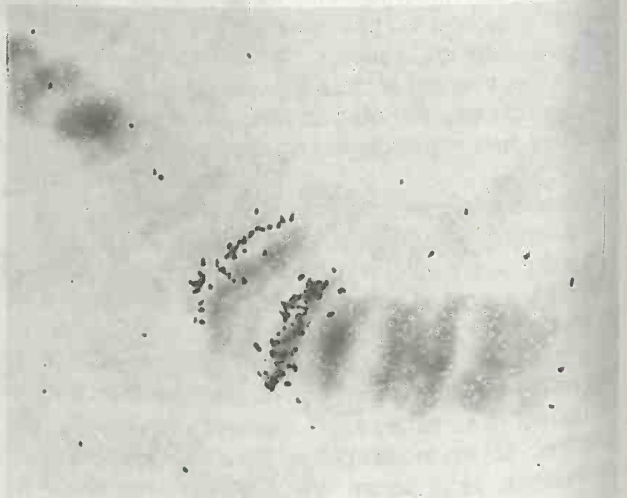


Figure 25.17

Une image agrandie des bandes 87A et 87C montre leur hybridation *in situ* avec l'ARN marqué extrait de cellules ayant subi un choc thermique.

Photographie aimablement fournie par Jose Bonner.



Les positions de gènes particuliers sur la carte cytologique peuvent être déterminées directement par la technique d'**hybridation *in situ*** ou **cytologique**, décrite au Chapitre 21, dans le cas de l'ADN hautement répétitif. Le protocole en est récapitulé dans la **Figure 25.17**.

Une sonde radioactive, représentant un gène (le plus souvent un clone marqué d'ADNc dérivé de l'ARNm) est hybridée *in situ* à l'ADN dénaturé des chromosomes polytènes. Par la superposition des grains sur une, ou des bandes particulières, l'autoradiographie permet d'identifier la position ou les positions de gènes correspondants. La **Figure 25.18** en montre un exemple. Avec ce type de technique à notre disposition, il est possible de déterminer directement dans quelle bande réside une séquence particulière.

LA TRANSCRIPTION BOULEVERSEE LA STRUCTURE DES CHROMOSOMES

Une des caractéristiques intéressantes des chromosomes polytènes est que les sites actifs peuvent être visualisés. Certaines des bandes sont trouvées par moment dans un état dilaté ou **boursoufflé** dans lequel le matériel chromosomique fait hernie par rapport à l'axe. La **Figure 25.19** montre, par exemple, de très grosses boursoufflures ou « puffs » (appelées anneaux de Balbiani).

Quelle est la nature du « puff » ? Il consiste en une région dans laquelle les fibres chromosomiques se déroulent par rapport à leur état habituel de condensation au sein de la bande. Les fibres restent en continuité avec celles de l'axe chromosomique. Les puffs proviennent habituellement de bandes uniques, bien que lorsqu'ils sont très gros, comme dans l'exemple des anneaux de Balbiani, le gonflement peut être tellement étendu qu'il cache le réseau de bandes sous-jacentes.

Le mode de distribution des puffs est lié à l'expression des gènes. Au cours du développement larvaire, les puffs apparaissent et régressent selon un mode défini, spécifique de tissu. On trouve un mode de distribution des puffs caractéristiques de chaque tissu, à un moment donné. Des puffs sont induits par l'hormone ecdysone, qui contrôle le développement de la *Drosophile*. Certains puffs sont induits directement par l'hormone ; certains sont induits indirectement par les produits des puffs précédents.

Les puffs sont des *endroits où s'effectuent les synthèses d'ARN*. Le point de vue généralement admis concernant la formation du puff est que l'expansion de la bande est une conséquence du relâchement de la conformation chromatinienne nécessaire à la synthèse de l'ARN. La

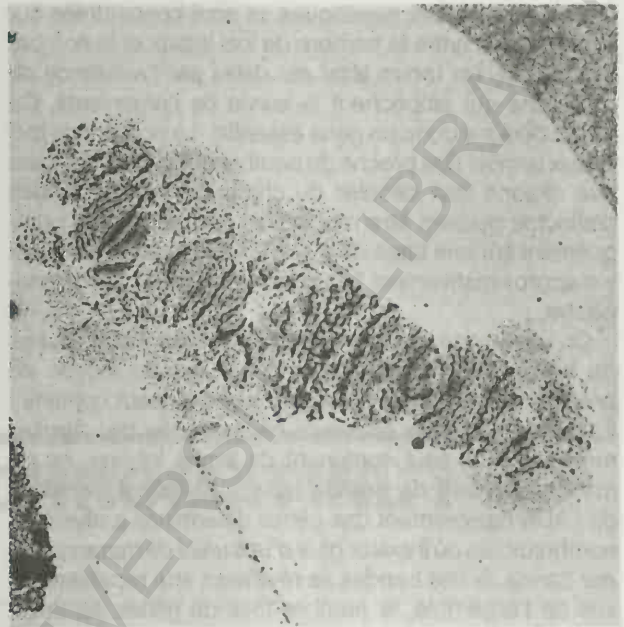


Figure 25.19

Le chromosome IV de *C. tentans* présente trois anneaux de Balbiani dans la glande salivaire.

Photographie aimablement fournie par Bertil Daneholt.

formation du puff a donc été considérée comme secondaire à la transcription. Un puff peut être produit par un seul gène actif.

Un cas inhabituel est fourni par un mutant qui présente un puff à un endroit qui synthétise habituellement des protéines de glue, bien que la mutation empêche la transcription des gènes situés sur cette bande. Ce mutant est la seule exception à la règle apparente qui veut que le puff reflète directement le niveau de la synthèse d'ARN.

Les sites où apparaissent les puffs diffèrent des bandes ordinaires par l'accumulation de protéines additionnelles. La caractérisation de ces protéines est actuellement seulement débutante. Nous savons qu'elles incluent l'ARN polymérase II et d'autres protéines associées à l'activité de transcription. Nous aimerions analyser l'ensemble complet des protéines qui s'accumulent dans les puffs, en particulier caractériser celles qui sont une cause plutôt qu'une conséquence de l'apparition des puffs. Alors il devrait être possible de déterminer la nature des événements moléculaires responsables de l'expansion du matériel génétique.

Ainsi, un puff représente une bande contenant un gène qui est transcrit activement. La bande est-elle une unité d'expression génique ? A cette question souvent débattue, il n'y a actuellement pas de réponse univoque.

Des expériences génétiques se sont concentrées sur la corrélation entre le nombre de loci létaux et le nombre de bandes. Un **locus léthal** est défini par l'existence de mutations qui empêchent la survie de l'organisme. Ce terme désigne donc un gène essentiel. Le nombre de loci létaux semble être proche du nombre de bandes ; et dans des régions bien définies du chromosome, les bandes distinctes peuvent être rapportées aux loci létaux, pratiquement sur une base de 1 sur 1. Ceci voudrait dire qu'il y a approximativement 5 000 loci létaux chez *D. melanogaster*.

Or, une bande moyenne contient environ 25 000 paires de bases d'ADN haploïde. Il n'y a pas de moyen de prédire *a priori* combien de gènes celle-ci peut contenir : il pourrait s'agir de plusieurs gènes serrés très étroitement, ou d'un seul contenant de larges introns, ou du matériel bordant de grande taille. Cependant, l'analyse de l'ADN représentant des gènes déterminés a révélé de nombreux cas où il existe *plus d'une unité de transcription par bande*. Si ces bandes se révélaient être représentatives de l'ensemble, le nombre total de gènes serait de plusieurs fois supérieur à 5 000. Nous ne savons pas encore si les différentes unités de transcription dans de telles bandes sont sujettes à des contrôles coordonnés. Il semble plausible qu'une expression suffisamment intense de chacun des gènes à l'intérieur d'une bande puisse être associée à l'apparition d'un puff.

Il y a donc, à présent, une certaine divergence entre les données génétiques et biochimiques. Les premières plaident pour l'existence probable d'un seul gène par

bande, les secondes plaident pour la présence de plusieurs gènes par bande. Il est possible de concilier ces données en considérant que de nombreux gènes peuvent être non-essentiels, au moins selon les critères des loci létaux. Ceci laisse en suspens le problème de la raison pour laquelle les loci létaux semblent distribués régulièrement, à raison d'environ un par bande.

Les caractéristiques présentées par les chromosomes en écouvillon et polytènes suggèrent une conclusion générale. Afin d'être transcrit, le matériel génétique est dispersé par rapport à son état habituel, beaucoup plus compact. La question que l'on garde à l'esprit est de savoir si cette dispersion au niveau macroscopique des chromosomes mime les événements qui surviennent au niveau moléculaire, à l'intérieur de l'amas de chromatine interphasique ordinaire.

POUR EN SAVOIR PLUS

La morphogenèse du phage a été détaillée dans *Gene Expression, 3, Plasmids and Phages* de Lewin (Wiley, New York 1978, p. 496-535, 642-673). Les chromosomes inhabituels discutés dans ce chapitre sont le sujet des pages 455-475 de *Gene Expression 2, Eucaryotic chromosomes*. Les centromères et les télomères ont été passés en revue par Blackburn & Szostak (*Ann. Rev. Biochem.* 53, 163-194, 1984) et par Clarke & Carbon (*Ann. Rev. Genet.* 19, 29-56, 1985). Les éléments fonctionnels requis par un chromosome ont été rassemblés pour faire un chromosome synthétique par Murray & Szostak (*Nature*, 305, 189-193, 1983).

CHAPITRE 26

STRUCTURE DE LA CHROMATINE : LE NUCLÉOSOME

La chromatine possède une organisation compacte au sein de laquelle la plupart des séquences d'ADN sont structurellement inaccessibles et fonctionnellement inactives. Une minorité de séquences actives sont situées à l'intérieur de cette masse. Quelle est la structure générale de la chromatine et quelle est la différence entre les séquences actives et inactives ?

La sous-unité fondamentale de la chromatine a le même type d'organisation chez tous les eucaryotes. Elle contient environ 200 pb d'ADN, dont la structure en forme de grains est maintenue par un octamère de petites protéines basiques. Les constituants protéiques forment un cœur interne ; l'ADN se trouve à la surface de la particule.

Au niveau de l'interaction ADN-protéine, la structure de la particule peut être définie en termes de parcours de l'ADN et de contacts entre l'acide nucléique et les protéines. On peut rechercher la liaison entre les différentes protéines individuelles.

On peut se demander si la conformation de la particule varie et dans quelle mesure de telles variations peuvent être impliquées dans les fonctions de la chromatine. Qu'advient-il de la particule quand la chromatine est transcrite ou répliquée ? Ces mécanismes peuvent nécessiter des protéines additives qui modulent le comportement de la particule.

Quand la molécule est répliquée, la série de particules doit être reproduite sur les deux molécules double brin

filles. Tout en se demandant comment s'effectue l'assemblage de la particule, nous devons rechercher ce qu'il advient des autres protéines présentes dans la chromatine. Comme la réplication rompt la structure de la chromatine, cela pose à la fois le problème du maintien de régions avec une structure spécifique tout en offrant une opportunité de changement de structure.

Le facteur de condensation général du matériel génétique suggère d'emblée que l'ADN ne peut être condensé directement dans la structure finale de la chromatine. Une hiérarchie au sein de l'organisation doit exister.

Le premier niveau est l'enroulement de l'ADN dans les « grains » avec un facteur de condensation d'environ 6. Ces particules sont des constituants invariables de l'hétérochromatine, de l'euchromatine et des chromosomes.

Le second niveau d'organisation est la torsion de cette série de grains en arrangement hélicoïdal constituant la fibre d'à peu près 30 nm que l'on retrouve à la fois dans la chromatine à l'interphase et les chromosomes en mitose (voir Figure 25.7). Dans la chromatine, le facteur de condensation de l'ADN est d'environ 40. On ne sait pas si la fibre est inchangée, ou si elle a un facteur de condensation accru dans le chromosome en mitose.

Le facteur final de condensation est déterminé par le troisième niveau d'organisation, l'empaquetage de la fibre elle-même. Il en résulte un facteur global de condensation d'au moins 1 000 pour l'euchromatine, interchangeable cycliquement avec la condensation dans les

chromosomes en mitose pour aboutir à un facteur global d'au moins 10 000. Ceci aussi est vraisemblablement une fonction modulée par des protéines accessoires, comme l'est la différence entre l'euchromatine et l'hétérochromatine.

LES CONSTITUANTS PROTÉIQUES DE LA CHROMATINE

La masse de la chromatine contient environ deux fois plus de protéines que d'ADN. Les protéines sont de deux types : les histones et les protéines non histones. L'ARN représente moins de 10 % de la masse d'ADN. La plus grande partie de l'ARN est formée de chaînes naissantes encore associées à leur matrice d'ADN.

Les **histones** sont les protéines les plus basiques dans la chromatine où leur masse est équivalente à celle de l'ADN. A l'origine, cinq classes d'histones ont été caractérisées en fonction de leurs proportions relatives de chaque type d'acide aminé basique, exprimées sous forme du rapport lysine/arginine. La désignation des histones porte encore la trace de cette description d'origine. Les mêmes classes peuvent être identifiées pratiquement chez tous les eucaryotes. Leurs propriétés sont regroupées dans le **Tableau 26.1**.

Les histones « riches en arginine » H3 et H4, et les histones « légèrement riches en lysine » H2A et H2B interagissent directement avec l'ADN pour former la série de premier niveau de particules dans la chromatine. H3 et H4 sont parmi les protéines connues les plus conservées au cours de l'évolution. Leurs séquences peuvent même être identiques dans des espèces aussi éloignées que la vache et le pois. Les types H2A et H2B sont retrouvés chez tous les eucaryotes, mais leur séquence présente une variation interspécifique non négligeable.

La constance de structure de l'histone suggère que les interactions histone-ADN, histone-histone et histone-protéines non histones peuvent être généralement similaires dans différentes espèces ; ainsi, nous pouvons déduire les règles générales qui gouvernent à la fois la formation de la particule primaire et le repliement d'une série de particules en structure d'ordre supérieur.

L'histone H1 « très riche en lysine » est impliquée dans les interactions entre les particules. Elle correspond en fait à un ensemble de plusieurs protéines dont les propriétés sont plutôt proches, avec des séquences en acides aminés qui se chevauchent. (Un variant du globule rouge sanguin aviaire est appelé **H5**.) Les histones H1 présentent des variations sensibles entre les tissus et les espèces (et cette classe est apparemment absente des levures). Bien qu'elles soient les histones les plus basiques, elles peuvent être extraites facilement de la chromatine sous l'action d'une solution saline diluée (0,5 M).

Comme leur nom barbare le suggère, les protéines **non histones** incluent toutes les autres protéines de la chromatine. Elles sont cependant soupçonnées d'être plus variables entre les tissus et les espèces, bien que l'étendue de cette variabilité n'ait pas encore été bien mise en évidence, et elles représentent une proportion plus faible de la masse que les histones. Elles contiennent aussi un plus grand nombre de protéines, si bien que chaque protéine est présente en quantité beaucoup plus faible que pour n'importe quelle histone.

Les protéines non histones ont des fonctions en rapport avec l'expression génique et la structure d'ordre supérieur. Ainsi, l'ARN polymérase peut être considérée comme une protéine non histone majeure. Les protéines HMG (high mobility group en anglais) comprennent une sous-classe distincte et bien définie de protéines non histones. L'étude des autres non histones se heurte au

Tableau 26.1

Les histones sont des protéines fortement basiques.

Histone	Acides Aminés Basiques		Acides Aminés Acides	Rapport Basique/Acide	Poids Moléculaire
	Lys	Arg			
H1	29 %	1 %	5 %	5,4	23 000
H2A	11 %	9 %	15 %	1,4	13 960
H2B	16 %	6 %	13 %	1,7	13 774
H3	10 %	13 %	13 %	1,8	15 342
H4	11 %	14 %	10 %	2,5	11 282

Les poids moléculaires indiqués sont ceux des histones bovines ; ils sont généralement similaires dans les autres espèces.

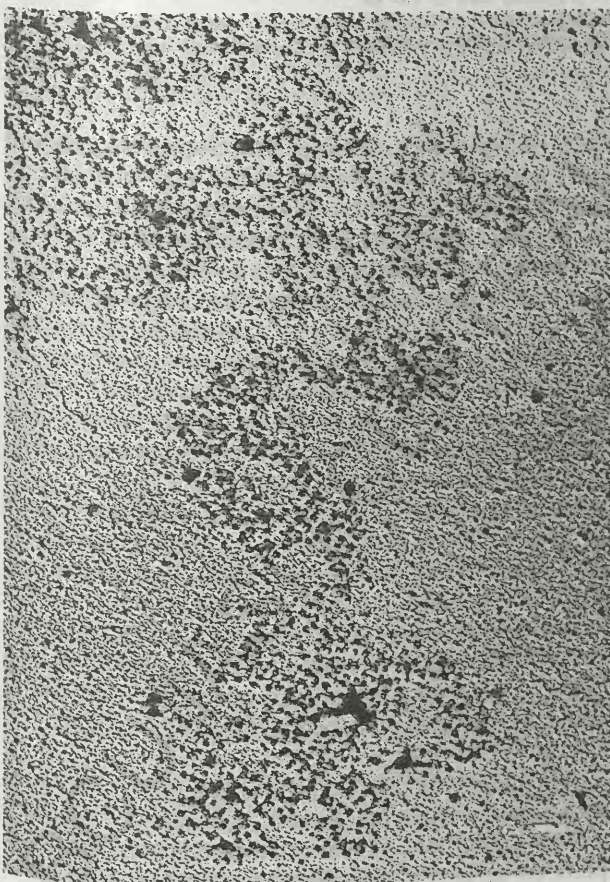
problème majeur de leur tendance à être contaminées par d'autres protéines nucléaires.

LE NUCLÉOSOME EST LA SOUS-UNITÉ DE BASE DE TOUTE LA CHROMATINE

Quand les noyaux en interphase sont mis en suspension dans une solution de faible force ionique, ils gonflent et éclatent en libérant des fibres de chromatine. La **Figure 26.1** montre un noyau lysé d'où les fibres s'échappent. A certains endroits, les fibres sont constituées de matériel étroitement condensé, alors qu'en d'autres endroits elles sont étirées, et peuvent être vues comme des particules distinctes appelées **nucléosomes**. Dans les régions par-

Figure 26.1
La chromatine qui s'échappe des noyaux lysés est constituée d'une suite de particules organisées d'une manière compacte. L'échelle est de 100 nm.

Photographie aimablement fournie par P. Chambon.



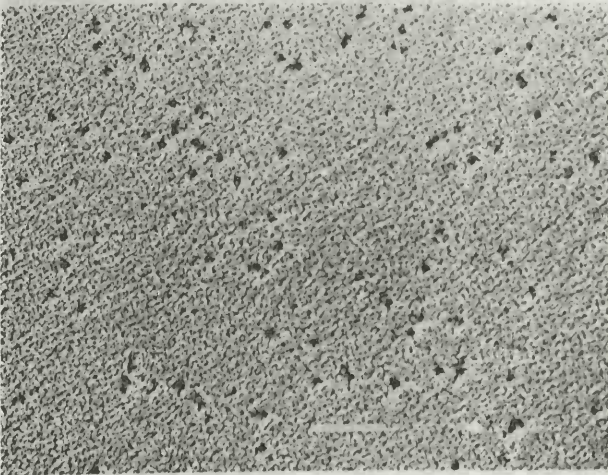
ticulièrement étirées, les nucléosomes individualisés sont connectés par un filament fin d'ADN libre double brin. Ainsi, un filament continu d'ADN double brin se faufile à travers cette série de particules.

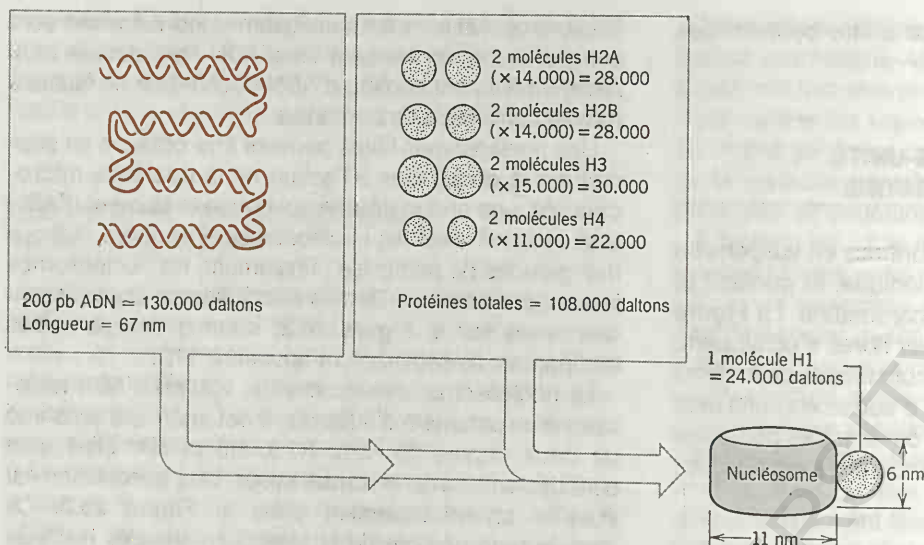
Des nucléosomes libres peuvent être obtenus en soumettant la chromatine à l'action de la **nucléase micrococcalle**, une endonucléase qui coupe le filament d'ADN à sa jonction avec les nucléosomes. Elle libère d'abord des groupes de particules ; finalement, les nucléosomes isolés sont obtenus. De tels nucléosomes apparaissent clairement sur la **Figure 26.2**, comme des particules compactes. Ils sédimentent environ à 11S.

Le nucléosome contient environ 200 pb d'ADN associées à un octamère d'histones. L'octamère est constitué de deux copies de H2A, H2B, H3 et H4. Elles sont connues en tant qu'**histones cœur**. Leur association est illustrée schématiquement dans la **Figure 26.3**. Ce modèle rend compte de la stœchiométrie des histones cœur dans la chromatine : H2A, H2B, H3 et H4 sont présentes en quantité équimolaire, avec une molécule de chaque par centaine de paires de bases de l'ADN. La très grande conservation d'H3 et d'H4 prouve que leur rôle peut être primordial et invariable, tandis que H2A et H2B apportent la variation spécifique d'espèces.

Une seule molécule d'H1 est associée à chaque nucléosome. Elle est présente dans la chromatine avec, globalement, une stœchiométrie de moitié inférieure à celles des autres histones. La totalité de H1 peut être enlevée

Figure 26.2
Les nucléosomes sont libérés par la digestion de la chromatine par la nucléase micrococcalle. L'échelle est de 100 nm.
Photographie aimablement fournie par P. Chambon.



**Figure 26.3**

Le nucléosome est constitué de masses d'ADN et d'histones (y compris H1) approximativement égales.

La masse théorique du nucléosome est de 262 000 daltons avec un rapport protéines/ADN de 1,0. La masse définie expérimentalement est généralement de l'ordre de 250 000-300 000 daltons, avec un rapport protéines/ADN d'environ 1,25. Toute protéine supplémentaire représente sans doute une faible quantité de protéines non histones associées aux nucléosomes.

sans affecter la structure du nucléosome, ce qui laisse penser que sa localisation est externe.

Quand la chromatine ou des noyaux isolés sont digérés par la nucléase micrococcale, l'ADN est coupé en fragments, dont la taille est un multiple d'une unité de longueur. La **Figure 26.4** montre « l'échelle » obtenue après fractionnement par électrophorèse sur gel. De telles échelles s'étendent sur une dizaine de bandes, voire plus, et l'unité de longueur déterminée par la distance entre les bandes successives est de l'ordre de 200 pb.

La **Figure 26.5** montre que l'échelle est constituée de groupes de nucléosomes individuels. Chaque fraction de nucléosomes engendre une bande d'ADN dont la taille correspond à une bande de l'échelle produite par la digestion de la chromatine. Le nucléosome monomérique contient une unité de longueur de l'ADN, le dimère contient deux unités de longueur de l'ADN et ainsi de suite.

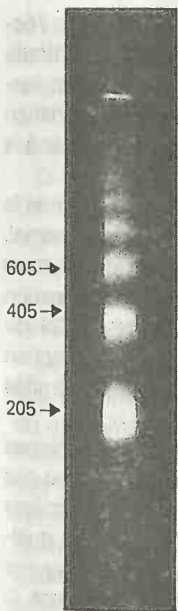
Ainsi, chaque « barre » de l'échelle représente l'ADN dérivé d'un nombre distinct de nucléosomes. On peut donc considérer l'existence de cette échelle de 200 pb dans toute la chromatine comme l'indication que l'ADN est organisé en nucléosomes. Ce type d'échelle micrococcale est obtenu quand seulement 2 % environ de

l'ADN nucléaire sont rendus acido-solubles (dégradés en petits fragments) par l'enzyme. Donc, une faible proportion de l'ADN est attaquée spécifiquement ; elle doit représenter des régions particulièrement sensibles.

Quand la chromatine s'étend en dehors du noyau, nous voyons souvent une série de nucléosomes reliés par un filament d'ADN libre (les perles d'un collier). Cette situation est-elle naturelle, avec les nucléosomes séparés par l'ADN libre, ou est-ce un artefact des conditions *in vitro* ? La nécessité d'une forte condensation de l'ADN *in vivo* suggère que, probablement, il y a peu (si ce n'est pas du tout) d'ADN libre habituellement.

Cette hypothèse est confirmée par le fait que plus de 90 % de l'ADN de la chromatine peuvent être récupérés sous forme d'une échelle de 200 pb. Par conséquent, presque tout l'ADN doit être organisé en nucléosomes. A l'état naturel, les nucléosomes sont sans doute étroitement condensés, l'ADN passant directement de l'un à l'autre. L'ADN libre est probablement engendré par la perte de quelques octamères d'histones au cours de l'isolement.

La longueur de l'ADN présent dans les nucléosomes peut varier quelque peu à partir de la valeur typique de 200 pb. Quand des génomes entiers sont caractérisés à partir de cellules particulières, une valeur moyenne bien

**Figure 26.4**

La nucléase micrococcale digère la chromatine des noyaux en une série multimérique de bandes d'ADN qui peuvent être séparées par électrophorèse sur gel.

Photographie aimablement fournie par Markus Noll.

définie est établie (plus ou moins 5 pb). Le plus souvent, cette moyenne se situe entre 180 et 200, mais des valeurs extrêmes aussi basses que 154 pb (dans un champignon) ou aussi élevées que 260 pb (dans le sperme d'oursin de mer) ont été trouvées.

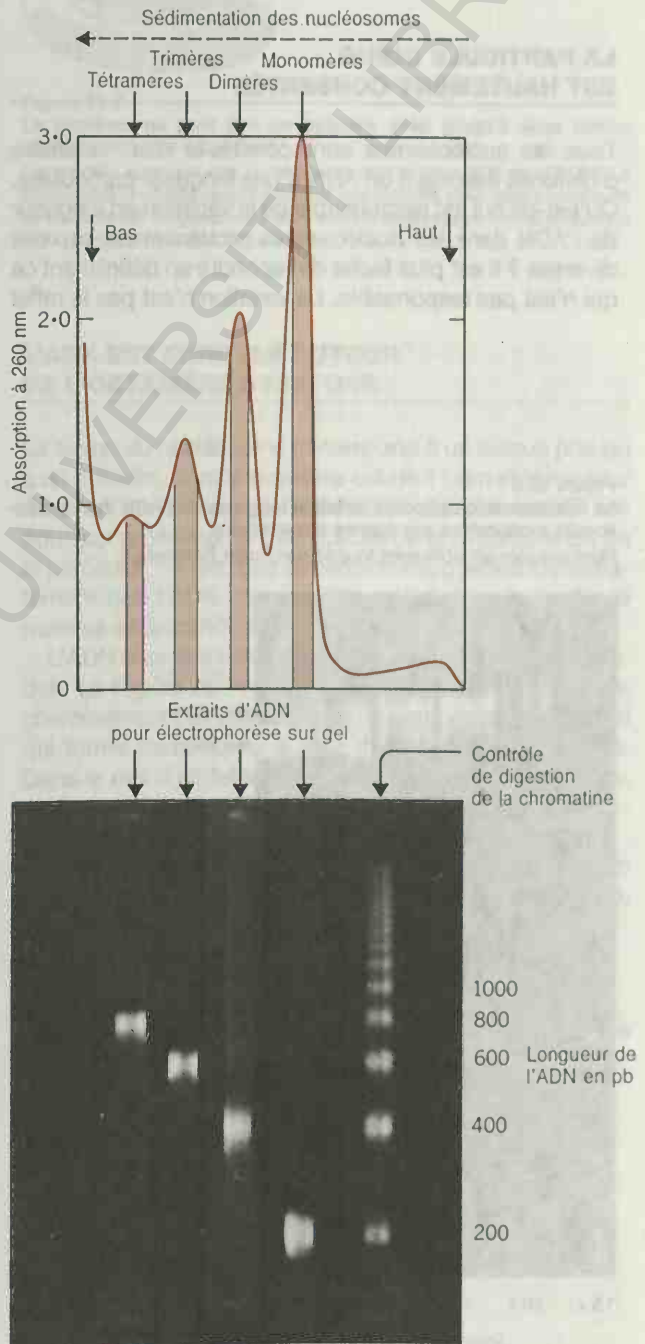
Cette valeur n'est pas nécessairement fixe. Elle peut varier au cours du développement embryonnaire. Dans le cas de l'oursin de mer, elle est réduite à une valeur plus typique au cours des divisions cellulaires de l'embryon précoce. La valeur moyenne peut être différente dans les divers tissus de l'organisme adulte. En outre, il

Figure 26.5

Chaque multimère de nucléosomes contient le nombre d'unités de longueur approprié.

Une préparation de nucléosomes a été fractionnée par sédimentation sur gradient de saccharose pour donner des monomères, dimères, trimères, etc... comme le montre la partie supérieure de la figure. Puis l'ADN de chaque fraction a été purifié et analysé par électrophorèse sur gel, comme le montre la partie inférieure de la figure. La partie droite de celle-ci représente en comparaison l'échelle électrophorétique obtenue après digestion de la chromatine.

Photographie aimablement fournie par John Finch.

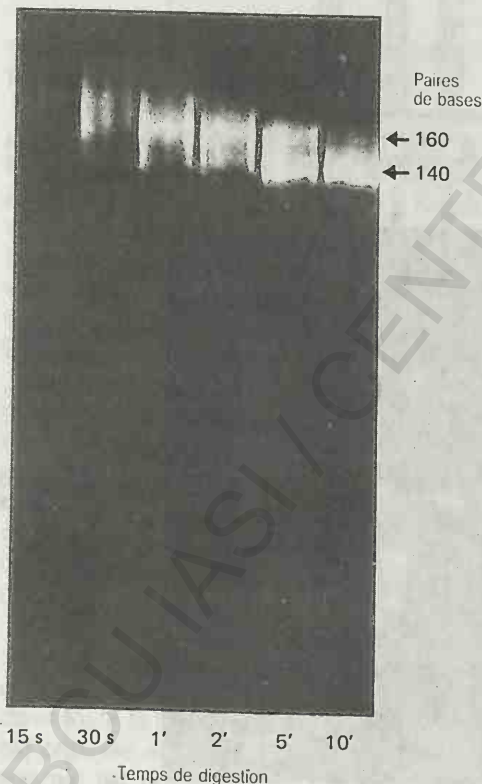


peut exister des différences entre les diverses parties du génome dans un unique type cellulaire ; des variations dans la moyenne génomique sont connues pour les ADN satellites et les gènes de l'ARN 5S.

LA PARTICULE CŒUR EST HAUTEMENT CONSERVÉE

Tous les nucléosomes sont constitués d'un octamère d'histones associé à un ADN d'une longueur particulière. Qu'est-ce qui est responsable de la variation en longueur de l'ADN dans les nucléosomes provenant de sources diverses ? Il est plus facile de répondre en définissant ce qui n'est *pas* responsable. La variation n'est pas le reflet

Figure 26.6
La nucléase micrococcalle réduit la longueur de l'ADN des nucléosomes monomères par étapes successives.
Photographie aimablement fournie par Roger Kornberg.



d'un changement dans l'association de l'ADN avec l'octamère d'histones, qui forme toujours une **particule cœur** contenant 146 pb d'ADN, quelle que soit la longueur totale de l'ADN dans le nucléosome. La variation de la longueur totale par nucléosome vient s'ajouter à la structure de base du cœur.

La particule cœur a été identifiée grâce à l'action de la nucléase micrococcalle sur le monomère nucléosomal. Initialement, l'enzyme coupe entre les nucléosomes, mais si son action peut se poursuivre au-delà de la formation des monomères, elle digère une partie de l'ADN du nucléosome isolé. Ceci s'effectue au cours d'une réaction dans laquelle l'ADN est « élagué » à partir des extrémités du nucléosome.

La longueur de l'ADN est réduite au cours d'étapes successives, comme le montre la **Figure 26.6**. Avec des noyaux de foie de rat, par exemple, les monomères nucléosomaux ont à l'origine 205 pb d'ADN. Puis quelques monomères apparaissent dans lesquels la longueur de l'ADN a été réduite à 160-170 pb. Finalement la longueur de l'ADN est réduite à celle de la particule cœur soit 146 pb. (Le cœur est assez stable, mais si la digestion est prolongée, d'autres coupures engendrent une **digestion limite**, dans laquelle les fragments les plus longs sont les 146 pb de l'ADN du cœur, tandis que les plus courts n'ont que 20 pb.)

Cette analyse suggère que l'ADN nucléosomal peut être divisé en deux régions. L'**ADN cœur**, d'une longueur constante de 146 pb, est relativement résistant à l'action des nucléases. L'**ADN de liaison (linker)** est constitué du reste de l'unité répétée. Sa longueur varie de 8 pb à 114 pb par nucléosome.

La nature bien définie de la bande d'ADN engendrée par le clivage initial de la nucléase micrococcalle suggère que la région disponible d'emblée pour cette enzyme est restreinte. Elle ne représente qu'une partie de chaque ADN de liaison. (Si la totalité de l'ADN de liaison était sensible, la bande se situerait à partir de 146 pb et au-delà de sa longueur répétée.) Mais dès la première coupure dans l'ADN de liaison, le reste de la région devient sensible et peut être retiré relativement rapidement par une action enzymatique ultérieure.

Les particules cœur ont des propriétés identiques à celles des nucléosomes eux-mêmes, bien qu'elles soient de plus petite taille. Leur forme et leur taille sont comparables à celles des nucléosomes, suggérant que la géométrie de base de la particule s'établit par des interactions entre l'ADN et l'octamère protéique dans la particule cœur. Les particules cœur étant obtenues plus rapidement sous forme d'une population homogène,

elles ont été utilisées pour une grande partie des études de structure de préférence aux préparations de nucléosomes. (Ces derniers ont tendance à varier du fait de la difficulté d'obtention d'une préparation dans laquelle les extrémités d'ADN n'ont pas été rabotées.)

Quelle est la nature du cœur et des régions de liaison ? Cette terminologie définit de façon opérationnelle les régions relativement plus ou moins sensibles à l'action de la nucléase. Cette description n'est pas une projection de leur structure réelle ; en particulier, elle n'implique pas que l'ADN de liaison ait une conformation plus étendue.

Le cheminement de l'ADN dans le nucléosome pourrait être continu, sans distinction évidente entre ces régions du monomère. Cependant, il s'agit d'une hypothèse de travail souvent établie afin de pouvoir extrapoler la structure du nucléosome à partir de la structure de la particule cœur. Il est aussi possible que le cheminement de l'ADN de liaison diffère de celui de l'ADN de la particule cœur, étant donné les importantes variations de longueur qu'ils présentent.

L'existence de l'ADN de liaison dépend de facteurs autres que les quatre histones cœur. Des expériences de reconstitution *in vitro* montrent que les histones sont capables intrinsèquement d'organiser l'ADN en particules cœur, mais ne forment pas de nucléosomes avec l'unité de longueur de l'ADN caractéristique de l'état *in vivo*. Le degré de superenroulement de l'ADN est un facteur important. L'histone H1 et/ou les protéines non histones peuvent influencer la longueur de l'ADN associé à l'octamère d'histones dans une suite naturelle de nucléosomes. « Les protéines d'assemblage » qui ne font pas partie de la structure de nucléosome peuvent être impliquées *in vivo* dans la formation des nucléosomes à partir d'histone et d'ADN (voir Chapitre 27).

L'histone H1 peut se situer dans la zone de jonction entre nucléosomes adjacents. Ceci serait en accord avec des résultats antérieurs montrant que H1 se détache le plus rapidement de la chromatine et que la chromatine plus rapidement « solubilisée » ; des résultats plus récents, allant dans le même sens, mettent en évidence qu'il est plus facile d'obtenir un collier de perles étiré sur un filament, une fois H1 détachée (voir plus loin).

H1 disparaît au cours de la dégradation des nucléosomes monomères. Elle peut être retenue sur les monomères qui ont encore 160-170 pb d'ADN ; mais elle est toujours perdue lors de la réduction finale à une particule de 146 pb. On peut en déduire que H1 serait vraisemblablement localisée dans la région de l'ADN de liaison immédiatement adjacente à l'ADN du cœur.

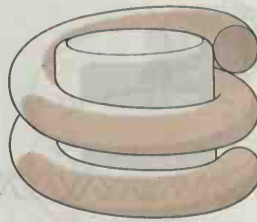


Figure 26.7

Le nucléosome peut être un cylindre avec jusqu'à deux spires d'ADN autour de sa surface.

Ce modèle suppose que le nucléosome a environ 200 pb d'ADN. La façon dont le cheminement pourrait être modifié pour tenir compte de la variation en longueur de l'ADN sur le nucléosome est un problème qui n'a pas encore été très étudié.

L'ADN EST ENROULÉ AUTOUR DE L'OCTAMÈRE D'HISTONE

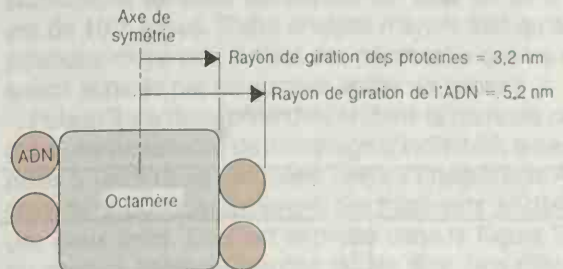
La forme du nucléosome correspond à un disque plat ou à un cylindre, dont le diamètre est de 11 nm et la hauteur de 6 nm. La longueur de l'ADN (200 pb) d'environ 67 nm équivaut grossièrement au double de la circonférence de la particule : 34 nm. Ces dimensions suggèrent immédiatement que l'ADN ne peut être aplati dans la particule mais se situe plutôt à l'extérieur.

L'ADN chemine autour de l'octamère de façon symétrique. La Figure 26.7 représente de façon schématique le cheminement de l'ADN comme un enroulement hélicoïdal qui forme deux spires autour du cylindre octamérique. Dans le cas d'un tel enroulement, les points d'« entrée » et de « sortie » de l'ADN sur le nucléosome sont relativement proches l'un de l'autre.

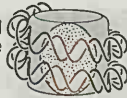
D'après ce modèle, si l'on considère une section verticale du nucléosome, comme l'indique la Figure 26.8,

Figure 26.8

Les deux tours d'ADN sur le nucléosome doivent être très proches l'un de l'autre.



Les sites de liaison à l'ADN sont proches l'un de l'autre sur le nucléosome



ADN restant



Sites de liaison



Protéine de liaison à l'ADN



Sur un ADN double brin étendu, les sites de liaison seraient très éloignés l'un de l'autre

Figure 26.9

Une protéine pourrait être en contact avec des séquences situées sur les différents tours qu'effectue l'ADN autour du nucléosome.

Que le point de départ de la transcription soit proche dans l'espace de la position - 80 est une possibilité qui est souvent considérée. Ainsi, ces deux points seraient simultanément en contact avec l'ARN polymérase (ceci pourrait représenter un exemple particulier du modèle moins détaillé illustré dans la Figure 9.16).

on s'aperçoit que les deux tours qu'effectue l'ADN sont proches. Il peut en découler une conséquence fonctionnelle. Puisqu'un tour de nucléosome nécessite environ 80 pb d'ADN, deux points séparés par 80 pb sur la double hélice libre peuvent être réellement rapprochés à la surface du nucléosome. Si, comme le montre la Figure 26.9, une protéine de liaison à l'ADN est simultanément en contact avec les deux tours d'ADN, les séquences qu'elle reconnaît dans l'ADN double brin pourraient être plus éloignées que ne le laisse supposer l'étendue apparente du polypeptide.

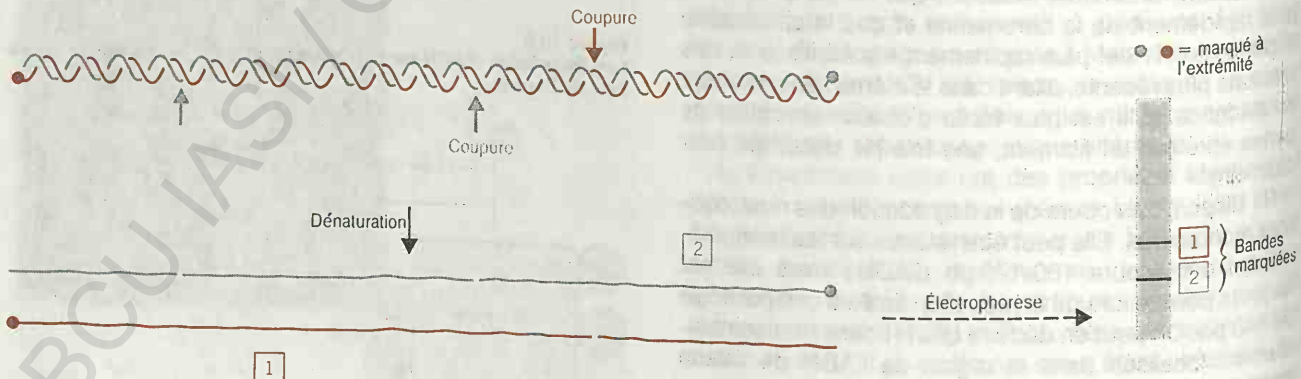
L'exposition de l'ADN à la surface du nucléosome explique pourquoi il est accessible au clivage par certaines nucléases. Les deux enzymes ADNase I et ADNase II

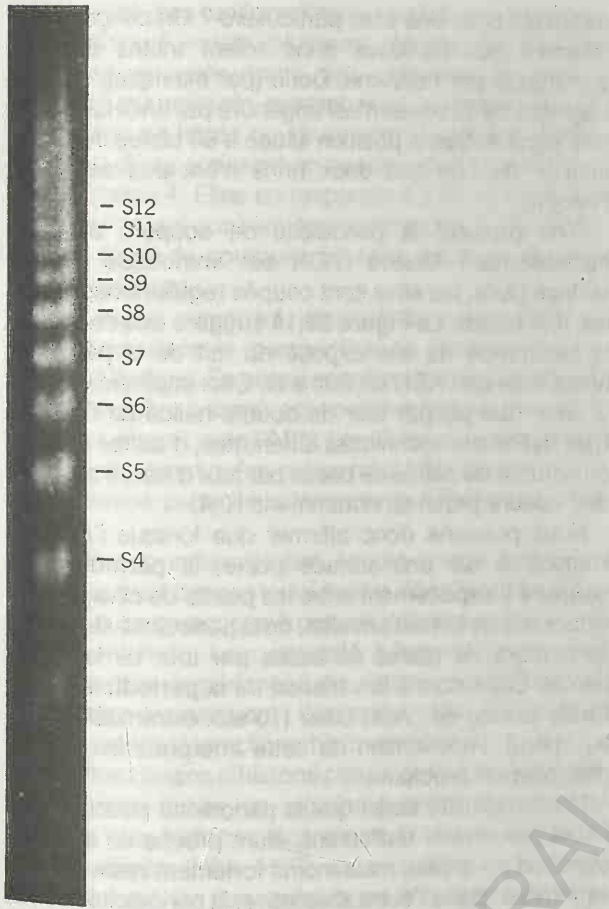
effectuent des coupures sur un brin d'ADN ; elles coupent une liaison sur un brin, sans affecter l'autre brin. Ainsi, il n'y a pas d'effet visible sur l'ADN double brin. Mais après dénaturation, des fragments plus courts que les monobrins de longueur totale sont libérés. Si les extrémités de l'ADN ont été marquées, les fragments terminaux seront identifiés par autoradiographie comme l'illustre la Figure 26.10. (Ceci est analogue à la technique de cartographie par restriction montrée dans la Figure 4.4.)

Quand l'ADN est libre en solution, il est coupé au hasard. L'ADN sur les nucléosomes peut aussi être coupé par les enzymes, *mais uniquement à intervalles réguliers*. Quand l'ADN est dénaturé puis soumis à une électrophorèse, une échelle est obtenue. Quand les sites de coupure

Figure 26.10

Les coupures dans l'ADN double-brin sont mises en évidence par les fragments obtenus après dénaturation de l'ADN en monobrins. Si l'ADN est marqué aux extrémités 5' (par exemple), seuls les fragments 5' sont visibles par autoradiographie. La taille du fragment définit la distance entre la coupure et l'extrémité marquée.



**Figure 26.11**

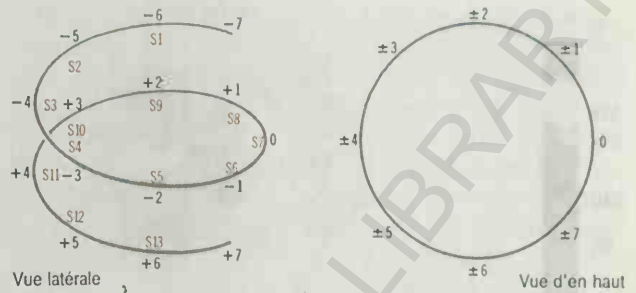
Les sites de coupe sont situés à des intervalles réguliers le long de l'ADN cœur, comme le montre une digestion des noyaux par l'ADNase I.

Photographie aimablement fournie par Leonard Lutter.

sont déterminés en utilisant un ADN dont les extrémités ont été marquées, une échelle du même type que celle qui apparaît dans la **Figure 26.11** est obtenue.

L'intervalle entre chaque bande de l'échelle équivaut approximativement à 10 bases. L'échelle s'étend sur la longueur totale de l'ADN cœur. Les sites de coupe sont numérotés de S1 à S13 (où S1 est à environ 10 bases de l'extrémité 5' marquée, S2 à environ 20 bases et ainsi de suite). Leurs positions par rapport à la superhélice sont illustrées **Figure 26.12**.

Tous les sites ne sont pas coupés avec la même fréquence : certains sont coupés plutôt efficacement, d'autres ne le sont pratiquement pas. Les enzymes

**Figure 26.12**

Deux systèmes de numérotation divisent l'ADN de la particule cœur en segments de 10 pb. Les sites peuvent être numérotés de S1 à S13 à partir d'une extrémité ; ou bien, si S7 sert à identifier la coordonnée 0 du centre de symétrie, ils peuvent être numérotés de -7 à +7.

ADNase I et ADNase II engendrent la même échelle, bien que les intensités des bandes soient quelque peu différentes. Ceci prouve que le mode de coupure représente une série particulière de cibles dans l'ADN, déterminée par son organisation, avec uniquement une légère préférence pour des sites particuliers imposés par l'enzyme.

La sensibilité de l'ADN nucléosomal aux nucléases est analogue à une expérience d'empreinte (footprinting). Ainsi, nous pouvons attribuer le défaut de réaction au niveau de sites cibles particuliers, à la structure du nucléosome, dans lequel certaines positions sur l'ADN sont rendues inaccessibles.

Quelle est la nature de ces sites cibles ? La **Figure 26.13** montre que chaque site est doté de 3-4 positions pouvant subir une coupure ; autrement dit, le site de coupure est défini à plus ou moins 2 pb. Ainsi, un site de coupure est constitué d'un court fragment d'ADN double brin, exposé à l'action de la nucléase sur 3-4 paires de bases. Les intensités relatives indiquent que certains sites sont préférés à d'autres.

A partir de ce diagramme, on peut calculer le point « moyen » qui est coupé. Aux extrémités de l'ADN, les paires de sites de S1 à S4 ou de S10 à S13 sont distantes l'une de l'autre de 10,0 bases chacune. Au centre de la particule, la distance qui sépare les sites de S4 à S10 est de 10,7 bases. (Cette analyse n'ayant trait qu'à des positions moyennes, il n'est pas nécessaire que les sites soient séparés par un nombre entier de bases.)

Puisqu'il y a deux brins d'ADN dans la particule cœur, dans une expérience de marquage d'extrémité, les extrémités 5' (ou 3') sont marquées, une sur chaque brin. Ainsi, le mode de coupure comprend des fragments qui dérivent des deux brins. Ceci est explicité dans la **Figure 26.10** où chaque fragment marqué dérive d'un brin différent.

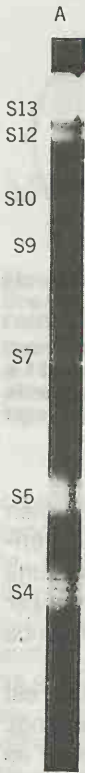


Figure 26.13

L'analyse de haute résolution montre que chaque site pour l'ADNase I est constitué de plusieurs liaisons phosphodiester adjacentes sensibles comme cela peut être vu dans cet exemple des sites S4 et S5 analysés dans les particules cœur marquées aux extrémités.

Photographie aimablement fournie par Leonard Lutter.

Le corollaire est que dans une expérience donnée, chaque bande marquée représente en fait deux fragments engendrés par une coupure à la même distance de l'une ou l'autre des extrémités marquées.

Comment pouvons-nous interpréter alors les affinités

distinctes pour des sites particuliers ? On conçoit généralement que les deux brins soient traités de façon symétrique par l'enzyme. Donc (par exemple), si aucun fragment de 80 bases n'est engendré par l'ADNase I, cela doit signifier que la position située à 80 bases de l'extrémité 5' de l'un des deux brins n'est pas sensible à l'enzyme.

D'où provient la périodicité de coupure de l'ADN nucléosomal ? Quand l'ADN est immobilisé sur une surface plane, les sites sont coupés régulièrement toutes les 10,6 bases. La **Figure 26.14** suggère que ceci reflète la récurrence du site exposé du fait de la périodicité hélicoïdale de l'ADN de forme B. Ceci implique qu'il doit y avoir 10,6 pb par tour de double hélice de l'ADN de type B. Par des techniques différentes, d'autres mesures du nombre de paires de bases par tour d'hélice ont fourni des valeurs proches, notamment 10,4.

Nous pouvons donc affirmer que lorsque l'ADN est immobilisé sur une surface plane, la **périodicité de coupure** (l'espacement entre les points de clivage) coïncide avec, de fait est un reflet, de la périodicité structurale (le nombre de paires de bases par tour de la double hélice). Étant donné la variation de la **périodicité structurale** le long de l'ADN cœur (10 aux extrémités et 10,7 au milieu), l'application de cette interprétation soulève d'emblée un problème.

Une possibilité serait que la périodicité structurale de l'ADN cœur varie réellement, étant proche de sa valeur standard au milieu, mais moins fortement resserrée aux extrémités. Dans l'autre alternative, la périodicité structurale de l'ADN serait constante mais la géométrie du nucléosome provoquerait un changement des longueurs moyennes séparant les coupures. Si l'approche de l'enzyme est limitée à certaines positions, elle pourrait opérer un glissement dans sa sélection entre les 3-4 liaisons disponibles. Ainsi, la périodicité de coupure n'est plus identique à la périodicité de structure.

La structure cristalline de la particule cœur suggère que l'ADN est organisé comme une superhélice plane, enroulée de 1,8 tours autour de l'octamère d'histone.

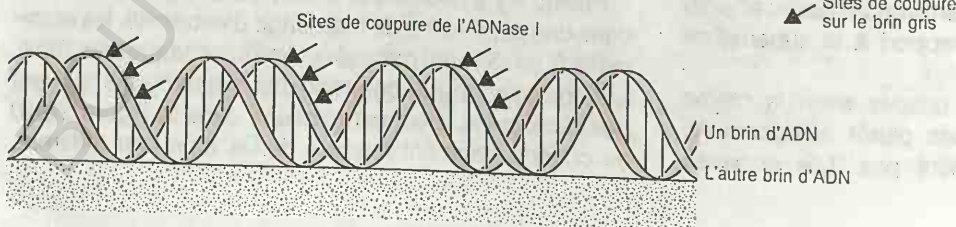


Figure 26.14

Les positions les plus exposées sur l'ADN réapparaissent avec une périodicité qui reflète la structure de la double hélice.

(Pour des raisons de clarté, les sites ne sont indiqués que sur un seul brin.)

L'ADN n'est pas uniformément courbé au sein de la superhélice ; il existe plusieurs régions de courbure accentuée, ou, sans doute, des « nœuds » discontinus, adjacents au point de contact avec le dimère H3 · H4. Les régions de haute courbure sont disposées symétriquement, et apparaissent en position plus ou moins 1 et plus ou moins 4. Elles correspondent à S6 et S8 et à S3 et S11, sites qui sont les moins sensibles à l'ADNase I. Le haut degré de courbure est sans doute responsable de ces variations.

Nous avons considéré jusque-là la construction du nucléosome dans la perspective de l'organisation de l'ADN à la surface. En ce qui concerne les protéines, nous devons chercher à savoir comment les histones interagissent entre elles et avec l'ADN. Les histones ne réagissent-elles correctement qu'en présence d'ADN ou bien possèdent-elles une capacité indépendante à former des octamères ?

Nous ne connaissons pas encore complètement les structures des histones individuelles dans le nucléosome, mais nous commençons à pouvoir déterminer leurs localisations relatives. Leur capacité à former des agrégats ainsi que des expériences de liaison croisée avec le nucléosome constituent l'essentiel des informations concernant les interactions histone-histone.

Les deux classes d'histone cœur forment des agrégats. Les histones riches en arginine forment un tétramère (H3₂H4₂). Les histones légèrement riches en lysine forment un dimère (H2A · H2B), qui a tendance à s'agréger ultérieurement. Un des agrégats pourrait être le tétramère (H2A₂H2B₂).

Des octamères d'histone intacts peuvent être obtenus soit par extraction à partir de la chromatine ou (plus difficilement) en laissant les histones s'associer *in vitro* dans des conditions de concentration élevée en sel et en protéines. L'octamère peut se dissocier en engendrant un hexamère d'histones ayant perdu le dimère H2A · H2B. L'autre dimère H2A · H2B est ensuite perdu séparément, laissant le tétramère H3₂H4₂. Ceci est en faveur d'une forme d'organisation dans laquelle le nucléosome peut avoir un « noyau » central constitué du tétramère H3₂H4₂, pouvant organiser *in vitro* l'ADN en particules qui présentent certaines des propriétés de la particule cœur.

Des études de cross linking (établissement de liaisons covalentes entre des protéines voisines) élargissent ces relations, montrant quelles paires d'histones sont l'une à côté de l'autre dans le nucléosome. (La difficulté avec de telles données est qu'habituellement, seule une faible proportion des histones peuvent être ainsi liées ; il est donc nécessaire d'être prudent en décidant si ce type de résultats représente les interactions majeures). A partir de ces données, un modèle a été élaboré pour l'organi-

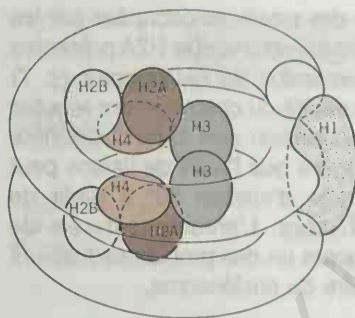


Figure 26.15

Dans un modèle symétrique du nucléosome, le tétramère H3₂H4₂ fournit le noyau de la structure.

Bien que la forme globale de l'octamère soit relativement bien définie, les histones individuelles sont représentées par des taches sans forme en raison du manque d'information les concernant.

sation du nucléosome. Il est décrit schématiquement dans la Figure 26.15.

Des études de structure ont montré que la forme générale de l'octamère d'histone isolé est semblable à celle de la particule cœur. Ceci laisse penser que la structure générale s'établit grâce aux interactions histone-histone. Les positions des différentes histones ont été attribuées à des régions de la structure octamérique sur la base de leur capacité à s'agréger et à répondre à la réaction de cross linking.

Le diamètre de l'octamère est déterminé par le tétramère H3₂H4₂. Les paires H2A · H2B s'adaptent comme deux dimères. L'ADN s'enroule deux fois autour de l'octamère. L'histone H1 pourrait « sceller » l'ADN dans le nucléosome en se liant à ses sites d'entrée et de sortie. Le modèle offre une double symétrie.

Une précision importante doit être faite en ce qui concerne le matériel utilisé au cours de ces expériences. Les données *in vitro* ont été obtenues à partir de particules cœur (du fait de la difficulté à obtenir une préparation de nucléosomes ayant une distribution homogène de taille d'ADN). Les modèles visant à concilier les différences dans le nombre de tours superhélicoïdaux considèrent que l'ADN de liaison suit le même cheminement que l'ADN cœur ; mais il pourrait être différent. Des hypothèses contradictoires à ce sujet ont été émises.

Quelques profils de digestion qui s'étendent au-delà de 200 pb après action de l'ADNase I indiquent que le même schéma peut s'étendre au-delà du cœur. Ceci indique que la même exposition périodique n'est pas seulement maintenue sur chaque nucléosome mais existe aussi entre les nucléosomes. En d'autres termes, l'ADN passe d'un nucléosome à un autre sans variation dans son cheminement, pour autant que la périodicité de coupure le laisse voir.

D'autre part, l'analyse des positions occupées par les histones suggère que les deux sous-unités H2A présentes à proximité des sites terminaux de l'ADN cœur (± 7) pourraient bloquer la poursuite du cheminement régulier de la superhélice. Leur extension vers le bord extérieur de la particule cœur suggère que l'ADN de liaison peut avoir à effectuer un virage d'environ 50° à partir de l'enroulement de l'ADN cœur. L'enroulement réel de l'ADN de liaison reste encore un des problèmes majeurs non résolus de la structure du nucléosome.

SUPERENROULEMENT ET PÉRIODICITÉ DE L'ADN

Le modèle de la Figure 26.7 montre les deux tours que l'ADN effectue autour du nucléosome. Si l'ADN n'était pas fixé par l'octamère d'histone, cet enroulement engendrerait moins de 2 tours superhélicoïdaux. Peut-on mesurer directement les degrés de superenroulement de l'ADN ?

Une grande partie des travaux sur la structure des ensembles de nucléosomes a été réalisée sur le virus SV40. L'ADN de SV40 est une molécule circulaire de 5 200 pb, soit environ 1 500 nm de circonférence. A la fois dans le virion et le noyau infecté, il est condensé dans une série de nucléosomes appelée **minichromosome**.

Isolé selon une technique habituelle, le minichromosome occupe un espace de 210 nm de long, ce qui correspond à un facteur de condensation d'à peu près 7 (pratiquement le même que celui du nucléosome lui-même, 67/11 égalent environ 6). Des changements de concentration saline peuvent le transformer en un « collier de perles » flexible dont le facteur de condensation globale est plus faible. Cela renforce l'idée qu'*in vitro*, les colliers de nucléosomes peuvent adopter, en fonction des conditions, plus d'une conformation.

Le superenroulement dans la chromatine peut être créé à plusieurs niveaux. D'abord, le superenroulement peut être la conséquence de l'enroulement de l'ADN autour du nucléosome. Il peut y avoir, ensuite, un superenroulement créé par les positions adoptées par les nucléosomes au sein des structures d'ordre supérieur. Certains de ces superenroulements peuvent être limités par des protéines. Des mesures directes de la densité de superenroulement montreront ainsi la tension moyenne de torsion de l'ADN qui résulte des superenroulements libres mais ne révéleront aucun des superenroulements limités de l'ADN.

Une façon de mesurer le superenroulement des nucléosomes du minichromosome est illustrée **Figure 26.16**. Les superenroulements libres du minichromosome lui-même sont d'abord relâchés afin que les nucléosomes forment un cercle de densité superhélicoïdale égale à 0. Puis les octamères d'histone sont extraits. Ceci permet

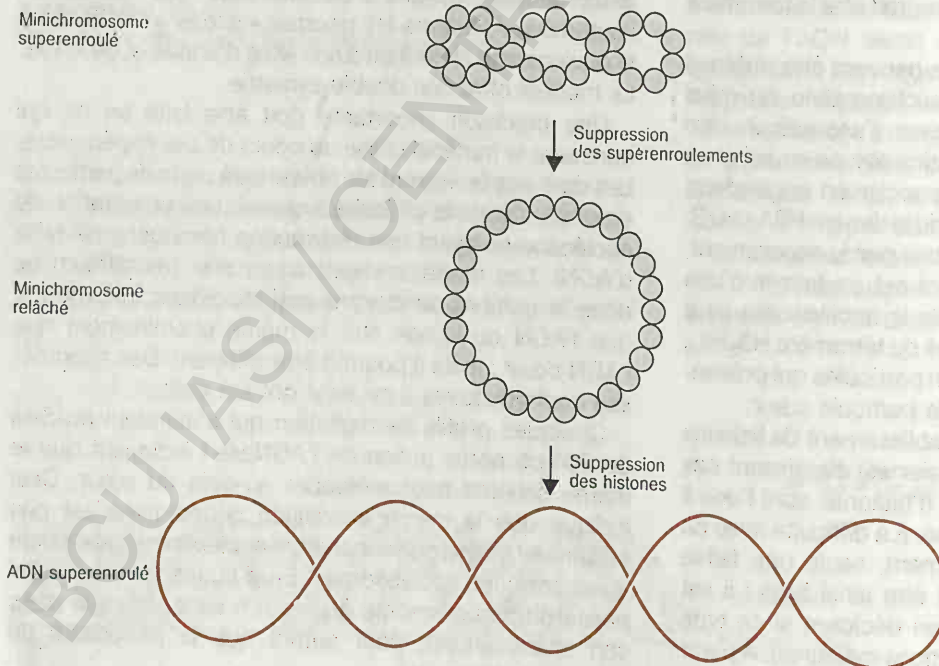


Figure 26.16
Les superenroulements du minichromosome peuvent être relâchés, engendrant alors une structure circulaire ; cette dernière, en perdant ses histones, engendre des superenroulements dans l'ADN libre.

à l'ADN de s'organiser librement. Chaque superenroulement qui était présent mais ne pouvait se manifester du fait des interactions ADN protéines dans le minichromosome apparaîtra dans l'ADN déprotéinisé comme moins 1 tour. On peut donc ainsi mesurer le nombre total de superenroulements dans l'ADN de SV40.

La valeur observée est proche du nombre de nucléosomes. On obtient un résultat inverse quand des nucléosomes sont associés *in vitro* sur un ADN de SV40 superenroulé. La formation de chaque nucléosome retire environ 1 superenroulement négatif.

Ainsi, le cheminement de l'ADN à la surface du nucléosome engendre à peu près 1 superenroulement négatif quand la protéine de liaison est absente. Comment peut-on expliquer l'écart entre la valeur mesurée de -1 tour par nucléosome et le modèle qui montre un cheminement de l'ADN équivalant à moins 2 tours superhélicoïdaux ?

On peut expliquer cette divergence en supposant que la périodicité de structure de l'ADN sur le nucléosome est différente de celle de l'ADN libre. Un modèle suggère que, de fait, il existe sur le nucléosome une périodicité de structure de 10,0 pb par tour qui est significativement plus faible que celle mesurée en solution (10,6).

Si cela est exact, quand l'ADN est libéré d'un nucléosome, sa structure deviendrait plus étroitement enroulée, puisqu'il y a par tour, un plus grand nombre de paires de bases. Cette variation réduirait le degré apparent de superenroulement. Supposons que sur le nucléosome le parcours de l'ADN soit équivalent à deux tours superhélicoïdaux. Puis l'octamère d'histone est libéré. Une partie de la contrainte torsionnelle va servir à accroître l'hélicité de l'ADN, le reste seulement pouvant être mesuré sous forme de superenroulement.

Ceci permet de réconcilier les modèles de moins 2 tours superhélicoïdaux par nucléosome avec les données qui identifient uniquement moins 1 tour superhélicoïdal, puisque la différence en périodicité (0,6 pb par tour d'hélice) multipliée par le nombre de tours d'hélice par nucléosome (supérieurs à 15) donne un résultat proche de 10 pb, ce qui correspond à un tour de la double hélice, absorbant ainsi potentiellement 1 tour superhélicoïdal négatif.

LES NUCLÉOSOMES SONT-ILS ORGANISÉS EN PHASE ?

Nous savons que les nucléosomes peuvent se reconstituer *in vitro* indépendamment de la séquence d'ADN, mais cela n'exclut pas la possibilité que leur formation

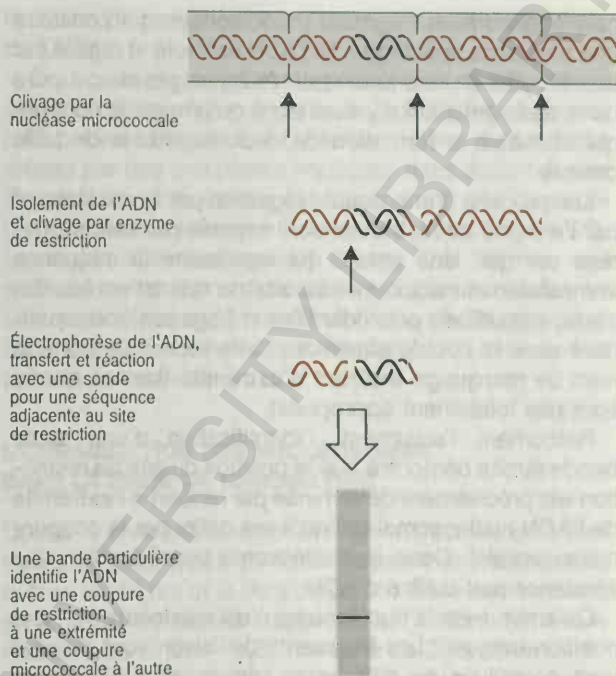


Figure 26.17

Les nucléosomes en phase placent les sites de restriction en des positions particulières par rapport aux sites de liaison clivés par la nucléase micrococcale.

soit contrôlée *in vivo* par un phénomène dépendant de la séquence. Une séquence particulière d'ADN est-elle toujours proche d'une certaine position *in vivo* par rapport à la topographie du nucléosome ? Ou bien les nucléosomes sont-ils disposés au hasard sur l'ADN, si bien qu'une séquence particulière peut apparaître localisée, par exemple, dans la région cœur d'une copie du génome et dans la région de liaison d'une autre copie ?

Pour essayer de répondre à cette question, il est nécessaire d'utiliser une séquence définie d'ADN ; plus précisément il faut déterminer la position d'un point défini de l'ADN par rapport au nucléosome. La Figure 26.17 illustre le principe de la procédure utilisée pour réaliser cette expérience.

Supposons que la séquence d'ADN soit organisée dans les nucléosomes selon une seule configuration particulière de telle sorte que chaque site sur l'ADN soit toujours localisé sur une position déterminée du nucléosome. Ce type d'organisation est appelé **nucléosome en phase**. Dans une série de nucléosomes en phase, les régions de liaison de l'ADN contiennent des sites particuliers.

Considérons les conséquences qui en découlent pour un seul nucléosome. Le clivage par la nucléase micrococ-

cale engendre un fragment monomérique qui constitue une séquence spécifique. Si l'ADN est isolé et digéré par une enzyme de restriction qui n'a qu'un site de coupure dans ce fragment, il n'y aura donc qu'une seule coupure qui aboutit à la formation de deux fragments de taille donnée.

Les produits d'une double digestion par la nucléase et par l'enzyme de restriction sont séparés par électrophorèse sur gel. Une sonde qui représente la séquence immédiatement adjacente au site de restriction (un des côtés) est utilisée pour identifier le fragment correspondant dans la double digestion. Cette technique porte le nom de **marquage indirect d'extrémité** (termes qui ne sont pas totalement appropriés).

Retournant l'argument, l'identification d'une seule bande étroite démontre que la position du site de restriction est précisément déterminée par rapport à l'extrémité de l'ADN nucléosomal (tel qu'il est défini par la coupure micrococcale). Donc le nucléosome possède bien une séquence particulière d'ADN.

Qu'arrive-t-il si le nucléosome n'est pas localisé en une position unique ? Les fragments de liaison sont maintenant constitués de différentes séquences d'ADN dans chaque copie du génome. Ainsi, à chaque fois, la position du site de restriction est différente ; en fait, toutes les localisations sont possibles par rapport aux extrémités de l'ADN nucléosomal monomérique. La **Figure 26.18** montre que la double digestion engendrera donc une large traînée allant des plus petits fragments détectables (environ 20 bases) jusqu'à la longueur du monomère d'ADN.

En décrivant ces expériences, nous avons considéré la nucléase micrococcale comme une enzyme qui coupe l'ADN dans les régions de liaison exposées, sans spécificité de séquence d'aucune sorte. Cependant, l'enzyme possède réellement une spécificité de séquence (basée sur la sélection de certaines séquences riches en A-T). Nous ne pouvons donc pas affirmer que l'existence d'une bande spécifique obtenue par la technique de marquage indirect d'extrémité représente la distance entre un site de coupure et la région de liaison. Elle pourrait tout aussi bien représenter la distance entre le site de coupure et un site préférentiel de la nucléase micrococcale !

Cette possibilité est vérifiée en traitant l'ADN nu exactement de la même façon que la chromatine. Si des sites préférentiels pour la nucléase micrococcale sont situés dans cette région particulière, des bandes spécifiques apparaîtront. Puis le profil de digestion pourra être comparé à celui qui est engendré à partir de la chromatine.

Une *différence* entre le profil de l'ADN de contrôle et celui de la chromatine apportera la preuve que les

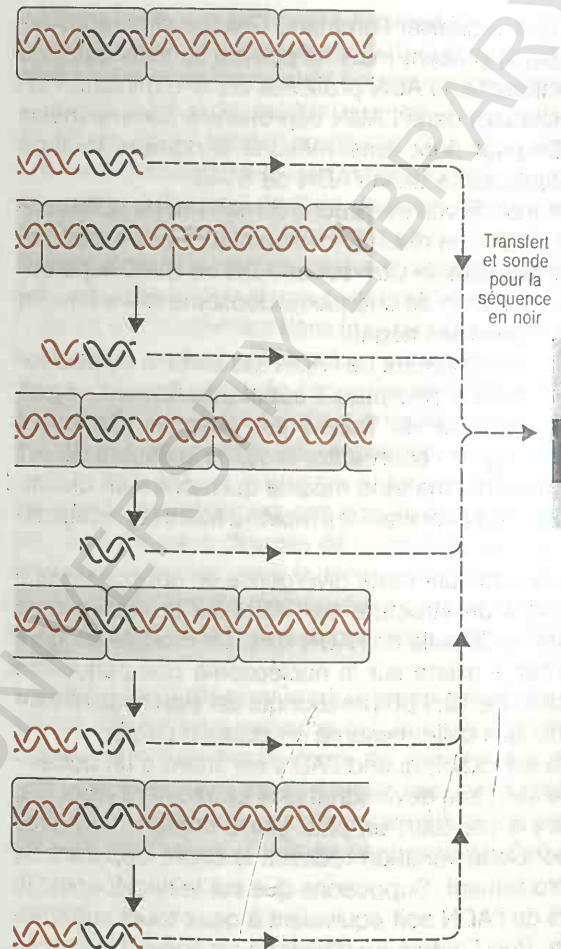


Figure 26.18

En l'absence de nucléosomes en phase, un site de restriction occupe toutes les localisations possibles dans les différentes copies d'un génome.

La figure montre le résultat de clivages à l'extrémité droite de la séquence en noir et aux jonctions entre les nucléosomes.

nucléosomes sont en phase : certaines bandes du contrôle de digestion peuvent disparaître dans la digestion du nucléosome quand des positions préférentiellement clivées ne sont pas disponibles ; de nouvelles bandes peuvent apparaître dans la digestion du nucléosome quand de nouveaux sites sont rendus préférentiellement accessibles par l'organisation nucléosomale.

L'analyse qu'illustrent les Figures 26.17 et 26.18 s'applique à une courte séquence unique. Cependant, pour de grands génomes, la précision n'est peut-être pas suffisante pour identifier sans difficultés un seul et unique fragment. Et nous voulons savoir si les nucléosomes sont

en phase avec une région substantielle d'ADN et non pas seulement avec une courte séquence particulière. D'où l'utilisation fréquente de séquences répétées en tandem au cours de l'étude des nucléosomes en phase. Cette approche, effectivement, amplifie le fragment qui est exploré : une seule sonde pourra détecter si la séquence correspondante est en phase dans toutes les unités répétées génomiques.

Des sondes ont été utilisées pour rechercher l'organisation en phase sur l'ADN satellite, l'ADNt, l'ADN 5S et les gènes d'histone. Dans de tels cas, l'organisation en phase du nucléosome n'apparaît que s'il existe une relation simple entre la longueur répétée nucléosomale et l'unité répétitive de la région en tandem. Ceci signifie qu'il doit exister un nombre entier de nucléosomes par répétition en tandem.

Dans la plupart des cas, les données expérimentales n'apportent pas un résultat aussi clair que ceux de la Figure 26.17, en faveur d'une disposition particulière des nucléosomes ; mais, parfois, ces résultats se rapprochent des bandes larges ou des traînées attendues pour une localisation au hasard. Ce que nous voyons, en fait, est un confinement des sites de coupure micrococcaux à un petit nombre de positions, relativement au site de restriction/défini. Ceci suggère que les nucléosomes ne peuvent peut-être adopter que quelques phases alternatives (2-4). Les nucléosomes sont parfois organisés en une seule phase, si bien que chaque nucléosome se trouve dans une position particulière.

L'organisation en phase du nucléosome pourrait s'accomplir au choix, de deux façons. Dans l'une, chaque nucléosome serait positionné spécifiquement en regard d'une séquence particulière d'ADN. Ceci n'est pas tout à fait en accord avec l'hypothèse du nucléosome en tant que sous-unité capable de se former entre toute séquence d'ADN et un octamère d'histone. Cependant, il existe des exemples où des changements dans la région modifient le profil de la phase, ce qui est en faveur d'une localisation particulière pour chaque nucléosome.

Dans l'autre alternative, il existe certaines séquences particulières qui sont utilisées préférentiellement pour assembler le premier nucléosome dans une région donnée ; puis les nucléosomes adjacents peuvent être assemblés séquentiellement avec une périodicité définie. (Si la construction des nucléosomes est soumise à certaines variations, par exemple si la longueur des fragments de liaison varie, disons de 10 pb, la spécificité de la localisation pourrait diminuer de façon continue, alors que l'on s'éloignera de la première localisation nucléosomale précisément définie.) Un point de départ dans la formation du nucléosome pourrait être défini par

la liaison de protéines non histones à un site spécifique de l'ADN.

Ce type d'organisation en phase peut résulter de la présence d'une région à partir de laquelle les nucléosomes peuvent être exclus. De telles régions peuvent être créées par des complexes impliqués dans le contrôle de l'expression génique ou qui engendrent une structure chromatinienne hautement ordonnée (voir Chapitre 27). La région d'exclusion peut former une *frontière* qui restreint les positions disponibles pour les nucléosomes adjacents. Nous serions en droit d'attendre que l'organisation en phase soit uniquement maintenue relativement près de la frontière.

L'ORGANISATION DES NUCLÉOSOMES DANS LA FIBRE DE CHROMATINE

Quand la chromatine est examinée en microscopie électronique, deux types de fibres peuvent être discernés : la fibre de 10 nm et la fibre de 30 nm. Elles sont décrites en fonction du diamètre approximatif du brin (en fait le diamètre de la fibre de 30 nm varie de 25 à 30 nm).

La **fibre de 10 nm** est essentiellement une suite continue de nucléosomes. Quelquefois, cependant, elle peut s'étirer, permettant ainsi de distinguer les nucléosomes comme les perles d'un collier, tel qu'il apparaît dans la Figure 26.19. La structure fibrillaire est obtenue dans des conditions de faible force ionique et en l'absence de l'histone H1. Ceci signifie que la formation de ces fibres est une propriété propre aux nucléosomes.

Comment les nucléosomes sont-ils organisés dans la fibre de 10 nm ? Si l'on imagine la particule comme une sorte de cylindre plat, les cylindres adjacents pourraient être disposés soit tranche contre tranche, soit face contre face. Ces types d'organisation peuvent être distingués par des techniques biophysiques telles que le balayage à neutrons ou le dichroïsme électrique, sensibles à l'orientation d'une sous-unité individuelle par rapport à l'axe de la fibre. Les résultats laissent supposer le type de modèle décrit par la Figure 26.20, dans lequel les cylindres sont orientés tranche contre tranche, les faces parallèles à l'axe de la fibre (ou tout au moins pas très inclinées). Les données impliquent que l'angle maximum entre les faces et l'axe peut être d'environ 20°.

Quand la chromatine est observée dans des conditions de plus grande force ionique et en présence d'histone H1, la **fibre de 30 nm** est visible. Un exemple en est donné Figure 26.21. La fibre présente une structure sous-jacente enroulée. Chaque tour comprend environ 6 nucléosomes, ce qui correspond à un facteur de condensation de 40 (c'est-à-dire que chaque μm le long de l'axe

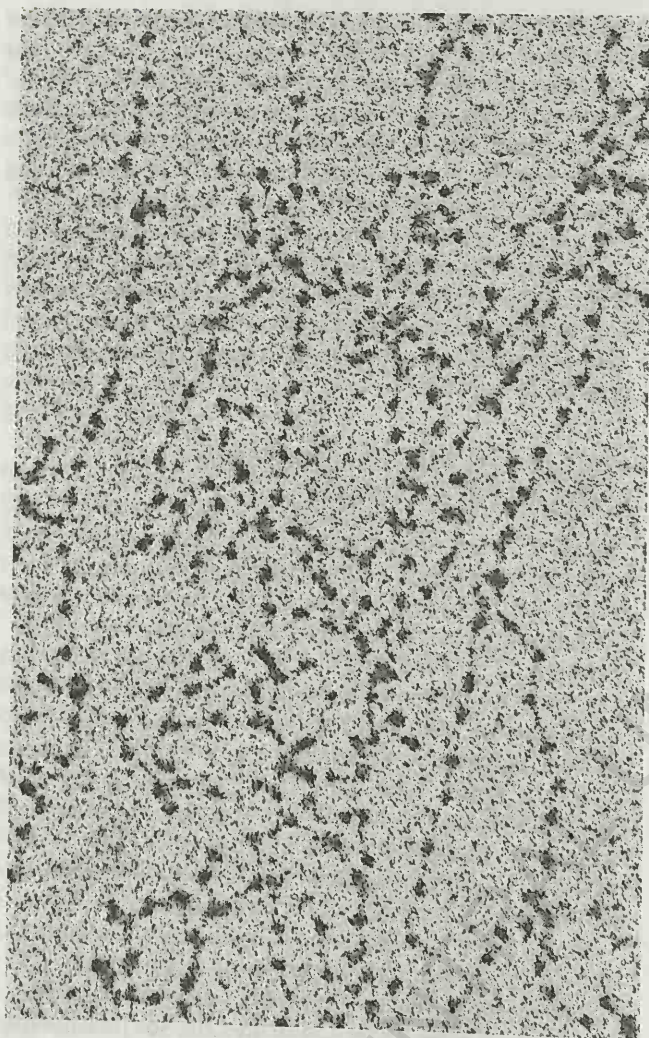


Figure 26.19
La fibre de 10 nm dans un état partiellement déroulé peut être vue sous la forme d'un chapelet de nucléosomes.
Photographie aimablement fournie par Barbara Hamkalo.

de la fibre contient 40 μm d'ADN). La fibre est le constituant de base à la fois de la chromatine en interphase et des chromosomes en mitose.

Un changement de force ionique permet la transformation réversible des fibres de 30 nm et de 10 nm. Ceci laisse supposer que le déploiement linéaire des nucléosomes dans la fibre de 10 nm est enroulé dans la structure de 30 nm si la force ionique est plus élevée et en présence d'H1.

Le solénoïde de la **Figure 26.22** représente le mode d'arrangement le plus probable des nucléosomes dans

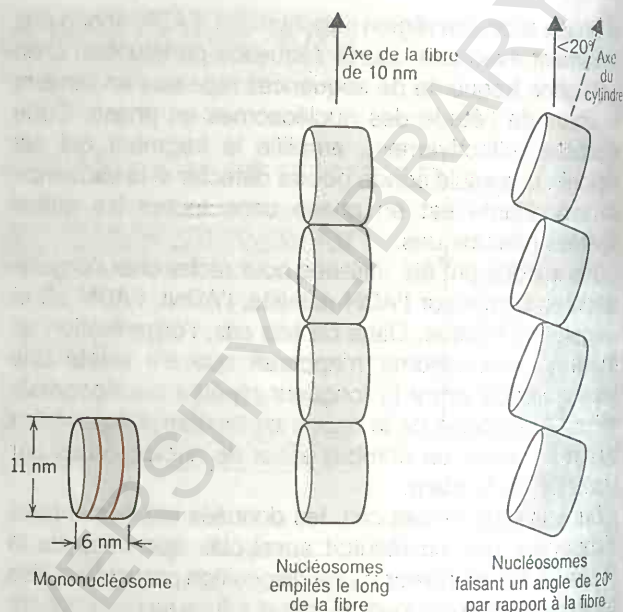


Figure 26.20

La fibre de 10 nm est constituée d'une série de nucléosomes organisés tranche sur tranche, entassés le long de la fibre ou inclinés de moins de 20° par rapport à l'axe

la fibre. Les nucléosomes s'enroulent de façon hélicoïdale, avec un angle d'environ 60° entre les faces des nucléosomes adjacents ; il y a 6 nucléosomes par tour.

Il est probable que les caractéristiques de la fibre de 30 nm ne soient pas totalement fixes mais puissent varier. Elles s'adaptent à la variation de longueur de l'ADN par nucléosome et permettraient aussi d'autres changements dans la compaction. Il n'est pas certain que la structure de la fibre soit identique dans la chromatine à l'interphase et dans les chromosomes en mitose.

Bien que la présence de H1 soit nécessaire à la formation de la fibre de 30 nm, les informations concernant sa localisation sont divergentes. La facilité relative avec laquelle elle est extraite de la chromatine est en faveur de sa présence à l'extérieur de l'axe de la fibre superhélicoïdale ; mais selon d'autres données, son accessibilité serait plus difficile à trouver dans les fibres de 30 nm que dans celles de 10 nm qui la retiennent, ce qui serait en faveur d'une localisation interne.

Comment de la fibre de 30 nm parvenons-nous aux structures spécifiques que présentent les chromosomes en mitose ? Et existe-t-il une spécificité ultérieure dans l'organisation de la chromatine à l'interphase ; des régions particulières des fibres de 30 nm jouissent-elles

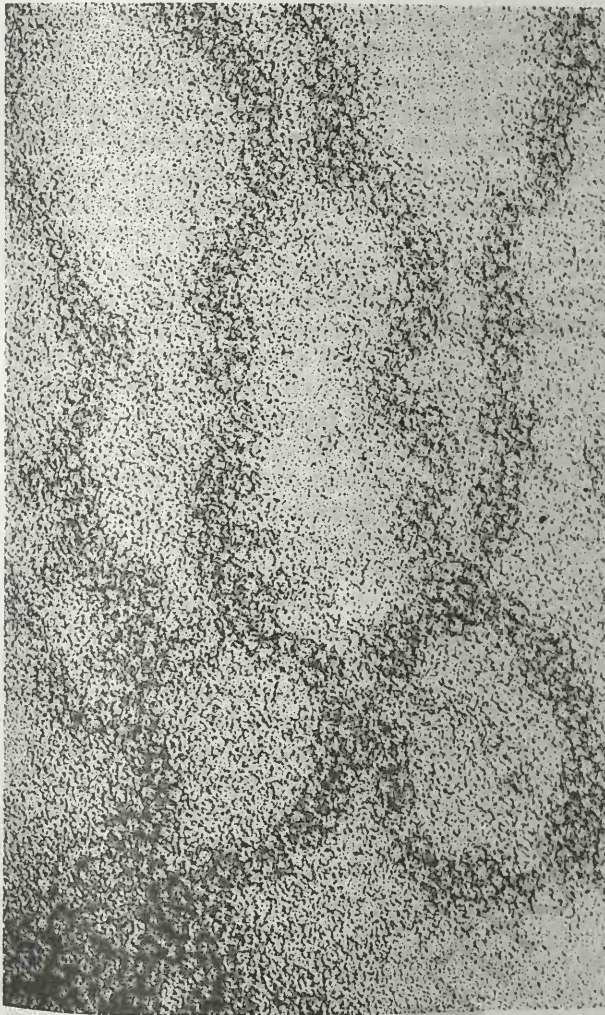


Figure 26.21
La structure de la fibre de 30 nm est enroulée.
La fibre est montrée avec le même grossissement que la fibre de 10 nm de la Figure 26.19.
Photographie aimablement fournie par Barbara Hamkalo.

d'une relation stable l'une par rapport à l'autre ou est-ce une organisation au hasard ? De telles questions restent encore aujourd'hui sans réponse.

BOUCLES, DOMAINES ET SQUELETTES

A l'interphase, la chromatine apparaît sous la forme d'une masse enchevêtrée qui occupe une grande partie du volume nucléaire, ce qui contraste avec l'ultrastructure

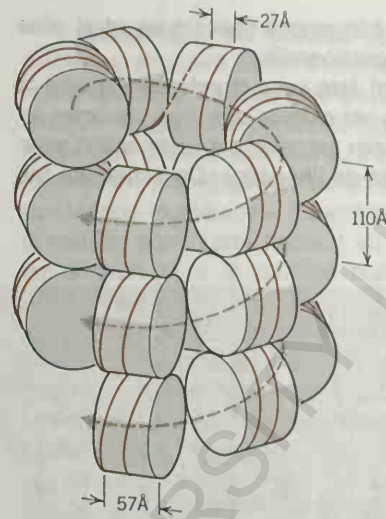


Figure 26.22
La fibre de 30 nm peut avoir un enroulement hélicoïdal de 6 nucléosomes par tour, organisé de façon radiale.

hautement organisée et reproductible des chromosomes en mitose. Quel élément contrôle la distribution de la chromatine interphasique dans le noyau ?

Une évidence indirecte de sa nature est apportée par l'isolement du génome sous la forme d'un seul constituant compact. En utilisant la même technique que celle décrite dans le Chapitre 25 pour l'isolement du nucléoïde bactérien, les noyaux peuvent être lysés sur un gradient de sucrose. Le génome ainsi libéré peut être récupéré par centrifugation. Ainsi isolé de *D. melanogaster*, il apparaît alors comme une fibre de 10 nm étroitement repliée, formée d'ADN et des 4 histones cœur.

Le superenroulement du corps compact peut être mesuré par sa réponse au bromure d'éthidium (voir Chapitre 25). Son niveau correspond à environ un superenroulement négatif toutes les 200 pb. Ces superenroulements peuvent être enlevés par la coupure par l'ADNase, bien que les nucléosomes restent présents. Ceci suggère que le superenroulement est dû à l'arrangement des nucléosomes. Il doit représenter la torsion à travers les jonctions des nucléosomes.

Un relâchement complet des superenroulements nécessite une coupure double brin toutes les 85 kb. Donc la longueur moyenne de l'ADN « fermé » (c'est-à-dire compris entre deux points immobilisés par des interactions avec des protéines) est d'environ 85 kb. Cette région serait constituée d'une boucle ou d'un domaine de nature identique à ceux déjà identifiés dans le génome bactérien. Il serait intéressant de savoir si ces boucles

correspondent à des séquences spécifiques et si elles ont une signification fonctionnelle.

Les boucles peuvent être vues directement quand la majorité des protéines est extraite des chromosomes en mitose. Les histones sont extraites par compétition avec le sulfate de dextran et de l'héparine. Ce traitement fait

également disparaître une grande partie des protéines non histones. Il en résulte un complexe d'ADN associé à environ 8 % du contenu protéique d'origine. Comme le montre la **Figure 26.33**, les chromosomes déplétés en histone forment un **squelette central** entouré d'un halo d'ADN.



Figure 26.23

Les chromosomes déplétés en histone sont constitués d'un squelette protéique sur lequel sont ancrées les boucles d'ADN.

Photographie aimablement fournie par Ulrich K. Laemmli.

Le squelette en métaphase est constitué d'un réseau dense de fibres. Les filaments d'ADN émanent du squelette, apparemment sous forme de boucles d'une longueur moyenne de 10-30 μm (30 à 90 kb). Si les boucles étaient rendues 40 fois plus compactes, de façon à s'ajuster dans une fibre de 30 nm, leur longueur moyenne serait de l'ordre de 0,25-1 μm , soit légèrement supérieure au diamètre du chromosome.

Les boucles peuvent être visualisées d'une autre façon. Quand les chromosomes sont débarrassés des cations divalents, des coupes transversales montrent les boucles sous forme de rangées de fibres de 10 nm, en étoile, d'une longueur moyenne de 3-4 μm . Ceci à nouveau est en faveur d'une organisation au sein de laquelle les boucles d'ADN d'environ 60 kb sont ancrées sur un squelette central de nature protéique.

L'ADN est-il uni à ce squelette via des séquences spécifiques ? Quand des squelettes isolés de *D. melanogaster* sont attaqués par des enzymes de restriction ou digérés par l'exonucléase III, des régions spécifiques de l'ADN restent liées au squelette. Le « blotting » (buvardage) de fragments du groupe de gènes d'histone répétés 100 fois montre que deux régions d'environ 200 pb de chaque répétition sont spécifiquement retenues par le squelette tandis que tout le reste de l'unité répétée est perdu. Ces deux fragments protégés sont séparés par une région sensible d'à peu près 100 pb. De ces résultats, on peut vraisemblablement déduire que chaque répétition dans le groupe de gènes forme une boucle séparée du squelette, ancrée par les deux séquences d'environ 200 pb*.

L'ADN peut être digéré sans que l'intégrité du squelette soit affectée, celui-ci étant constitué d'un ensemble de protéines non histones. L'une de ces protéines est peut-être la topo-isomérase II, ce qui peut signifier que le squelette peut effectuer des changements topologiques dans l'ADN qui lui est lié.

A l'interphase, les cellules possèdent une *matrice nucléaire*, structure filamenteuse à l'intérieur de la membrane nucléaire. La chromatine apparaît souvent liée à la matrice ce qui, selon certaines hypothèses, pourrait être nécessaire à la transcription ou à la réplication, la matrice fournissant probablement les activités nécessaires, telle que la topo-isomérase II. Il serait intéressant de connaître la relation entre l'attachement au squelette des cellules en division et à la matrice nucléaire des cellules à l'interphase.

L'apparence du squelette est proche de celle des paires de chromatides sœurs en mitose. Les squelettes des chromatides sœurs sont habituellement fortement reliés, mais peuvent quelquefois être séparés, associés seulement par quelques fibres. Cette structure pourrait-elle être responsable du maintien de la forme des chromosomes en mitose ? Pourrait-elle être engendrée en regroupant les constituants protéiques qui fixent habituellement les bases des boucles dans la chromatine à l'interphase ?

POUR EN SAVOIR PLUS

Les connaissances accumulées sur le nucléosome peuvent être reprises depuis l'esquisse de mise en évidence à partir de laquelle **Kornberg**, à l'origine, proposa un modèle (*Science* **184**, 868-871, 1974) jusqu'à l'apport massif de preuves maintenant assemblées dans les revues telles que celle de **Kornberg** (*Ann. Rev. Biochem.* **46**, 931-954, 1977), **McGhee & Felsenfeld** (*Ann. Rev. Biochem.* **49**, 1115-1156, 1980), et **Lewin** dans *Gene Expression, 2, Eucaryotic Chromosomes* (Wiley, New York, 1980, p. 332-393). Les enroulements de l'ADN et les nucléosomes ont été succinctement analysés par **Wang** (*Cell* **29**, 724-726, 1982) et **Felsenfeld & McGhee** (*Cell* **44**, 375-377, 1986). Un modèle de l'octamère d'histone a été développé par **Richmond et al.** (*Nature* **311**, 532-537, 1984), qui fait aussi état des analyses de structures antérieures.

* N. d. T. : Chez *D. melanogaster*, rappelons (voir Chapitre 19, p. 412, Figure 19.10) que chaque unité de répétition de gènes d'histones a 4,8 kb et code pour toutes les histones. Il semblerait donc ici, qu'existe une disposition en boucles bien plus petites que les 60 kb indiqués plus haut.

CHAPITRE 27

LA NATURE DE LA CHROMATINE ACTIVE

Décrire la chromatine comme un filament d'ADN double brin enroulé autour d'une série de nucléosomes est une première étape cruciale vers la visualisation de l'état du matériel génétique dans le noyau. Cette vue quelque peu statique justifie la structure de la sous-unité individuelle et (dans une certaine mesure) sa relation avec la sous-unité suivante. Cependant, la *flexibilité* de l'organisation des nucléosomes doit être suffisante pour satisfaire les diverses sollicitations, structurale ou fonctionnelle, auxquelles est soumise la chromatine.

Les variations cycliques de la condensation affectent la masse entière de l'euchromatine. Au cours de la division cellulaire, l'euchromatine doit devenir plus fortement condensée dans les chromosomes en mitose. Cette transition est sans doute contrôlée par des changements dans les protéines qui sont largement distribuées dans la chromatine.

Réplication et transcription sont des phénomènes locaux qui, vraisemblablement, nécessitent une dispersion de structure. La réplication qui se déroule suivant une série d'événements individuels, localisée dans les réplicons, engendre des régions d'ADN dupliqué double-brin, chacune associée à un ensemble d'octamères d'histone. Les phénomènes qui interviennent dans la reproduction du nucléosome ne sont pas encore définis. Nous aimerions connaître ce qui arrive au nucléosome au cours de la réplication et comment les nouveaux nucléosomes sont assemblés.

De toute façon, la séparation des brins parentaux d'ADN semble inévitablement briser la structure de la fibre de 30 nm et sans doute aussi celle de la fibre de 10 nm. Il serait intéressant de connaître l'étendue de cette dislocation. Est-elle limitée à la proximité immédiate du point de synthèse de l'ADN ou s'étend-elle au delà ? Existe-t-il des différences de structure discernables entre les régions qui se sont répliquées et celles qui doivent l'être ? La difficulté majeure dans l'analyse de la structure d'une région particulière en cours de réplication réside dans le caractère transitoire de ce phénomène.

La transcription sous-entend également le déroulement de l'ADN et par là même nécessite sans doute le déploiement de la fibre dans des régions limitées de la chromatine. Une vue simpliste suggère que ce procédé a besoin d'avoir « les coudées franches ». Les caractéristiques des chromosomes polytènes et en écouvillon décrits dans le Chapitre 25 indiquent qu'une organisation structurale plus large peut être associée à l'expression génique.

Nous aimerions connaître les changements de structure qui interviennent quand un gène est transcrit. Est-ce la structure globale de cette région qui varie ? La séquence transcrite reste-t-elle sous forme de nucléosomes ? S'il en est ainsi que deviennent-ils quand l'ARN polymérase transcrit l'ADN ? Qui assure l'accessibilité initiale du promoteur à l'enzyme ?

Pouvons-nous identifier des ensembles de nucléoso-

mes dont les différentes propriétés expliquent la structure ou la fonction de régions particulières ? En outre, les nucléosomes représentent-ils le seul type de structure protéines-ADN dans le filament double brin ou bien d'autres structures sont-elles présentes pour délimiter des sites particuliers ?

Quelle que soit l'importance de ces questions, elles constituent réellement un préliminaire à la conception de l'expression génique. *Quel est l'élément qui modifie l'état d'un gène de façon à ce qu'il soit transcrit au bon moment et au bon endroit ?* A l'opposé, quand et comment les gènes sont-ils éteints ?

ASSEMBLAGE DES NUCLÉOSOMES CONTRE REPRODUCTION DE LA CHROMATINE

La reproduction de la chromatine n'implique pas de période prolongée durant laquelle l'ADN est dépourvu de nucléosomes. Une fois l'ADN répliqué, les nucléosomes sont rapidement engendrés sur les deux nouveaux brins. Ceci est illustré par la photographie en microscopie électronique de la **Figure 27.1**, qui montre un fragment d'ADN nouvellement répliqué dont les deux segments fils double brin sont déjà recouverts de nucléosomes.

Comment les histones s'associent-elles à l'ADN pour créer les nucléosomes ? Cette question déroutante a longtemps été débattue. Les histones *forment-elles à l'avance* une protéine octamère autour de laquelle l'ADN s'enroule ultérieurement ? Ou bien l'ADN se lie-t-il à un noyau H3·H4, auquel viennent ensuite s'ajouter les dimères H2A·H2B ?

L'auto-assemblage *in vitro* est un processus lent, limité par la tendance des particules qui s'assemblent à précipiter. Il est difficile de reproduire les conditions physiologiques. La **Figure 27.2** illustre les différents moyens qui ont été utilisés pour assembler les nucléosomes *in vitro*.

Les essais de production des nucléosomes *in vitro* reposent essentiellement sur le processus qui met en présence l'ADN libre et les histones. Mais ce qui a réellement lieu *in vivo* c'est la reproduction de la chromatine. Une portion d'ADN déjà associée aux nucléosomes est répliquée, donnant naissance à deux doubles brins fils. A ce point précis, qu'advient-il des nucléosomes préexistants ? Les octamères d'histones se dissocient-ils en histones libres pour être réutilisés ou demeurent-ils assemblés ?

Des expériences de liaison croisée tendent à montrer que l'octamère d'histone est **conservé**, survivant en tant que tel au cours des cycles de réplication. Les « anciens » octamères s'associent-ils d'une façon particulière aux

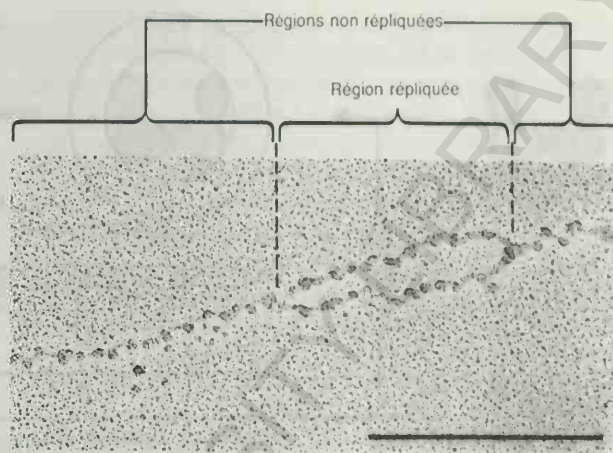


Figure 27.1

Les régions répliquées de la chromatine contiennent des nucléosomes sur les deux brins fils de l'ADN.

Photographie aimablement fournie par Steven L. McKnight.

ADN dupliqués ? La **Figure 27.3** montre les conséquences d'une réplication conservative ou dispersée.

Dans une réplication conservative, les octamères préexistants resteraient avec un des doubles brins fils alors que les « nouveaux » octamères s'assembleraient sur l'autre. Dans une réplication dispersée, les « anciens » octamères seraient dispersés à la fois sur les doubles brins fils. Les essais de distinction entre ces deux modèles ont nécessité l'utilisation de conditions inhibitrices de la production de nouvelles histones (inhibition de la synthèse protéique par la cycloheximide). On s'aperçoit alors que les anciens octamères se dispersent au hasard sur les deux brins fils.

Quand la chromatine se reproduit, les nucléosomes peuvent par conséquent avoir une double provenance. Les octamères d'histone existants sont déplacés de l'ADN pour lui permettre de se répliquer. Ils sont conservés et peuvent se réassocier à l'un ou l'autre des doubles brins fils. Cependant un nombre égal d'octamères doit être formé à partir des histones nouvellement synthétisées. Nous ne savons pas encore si ces octamères sont assemblés avant de s'associer à l'ADN ou s'ils suivent une voie alternative en s'assemblant sur l'ADN.

La structure de la fourche de réplication apparaît différente de celle des autres régions. Elle est plus résistante à la nucléase micrococcalle et elle est digérée en bandes dont la taille diffère de celle de l'ADN nucléosomal. Ceci suggère qu'un grand complexe protéique est engagé dans la réplication de l'ADN, mais que les

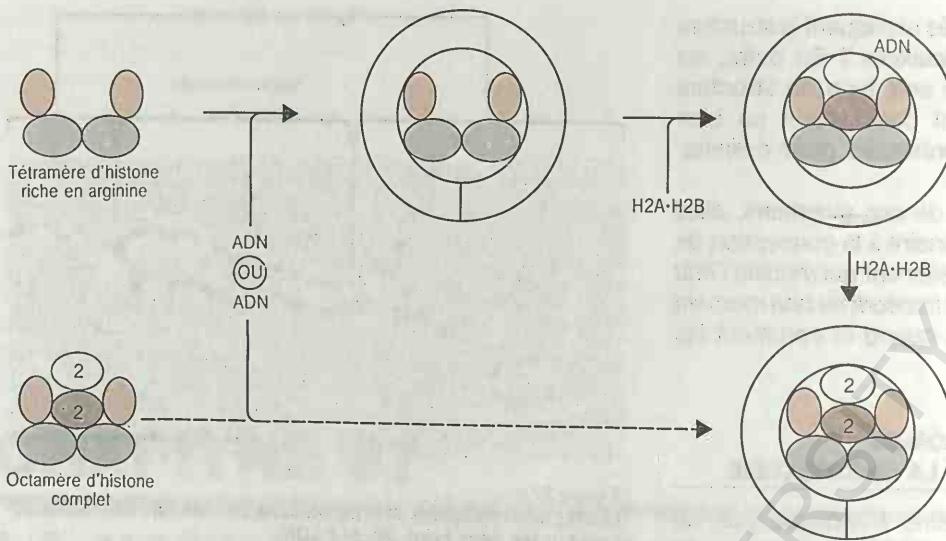


Figure 27.2

In vitro, l'ADN peut soit interagir directement avec un octamère d'histone intact (lié de façon croisée) soit s'assembler avec le tétramère $H3_2H4_2$ auquel deux dimères $H2A \cdot H2B$ sont ajoutés.

nucléosomes se reforment plus ou moins immédiatement derrière lui, alors qu'il progresse le long du brin.

Outre les histones nucléosomales, la chromatine contient également des protéines non histones qui doivent elles aussi être reproduites. Puisque le complément non histone varie vraisemblablement avec le phénotype de la cellule, sa reproduction implique le maintien de caractéristiques cellulaires spécifiques. Ainsi, la signification de l'existence éventuelle de modèles de ségrégation pour les protéines au cours de la réplication de l'ADN s'étend au-delà de l'assemblage des nucléosomes. Une des principales questions à laquelle nous aimerions répondre est celle de la perpétuation des états alternés de la structure chromatinienne au cours de la division cellulaire.

Considérons un gène qui est activé (ou réprimé) par la liaison à l'ADN d'une certaine protéine spécifique régulatrice et/ou par une modification dans la structure de la chromatine. Comment cet état particulier est-il transmis aux chromosomes dupliqués ?

Si toutes les protéines se dissocient de l'ADN pendant la réplication, cet état spécifique doit être rétabli à chaque cycle cellulaire. Cependant un modèle de ségrégation pourrait être utilisé pour propager l'information à propos de l'état d'expression du gène. Il est possible qu'une structure spécifique puisse être *perpétuée* par ségrégation et duplication pendant la réplication de l'ADN. Ainsi un complexe de protéines non histones pourrait s'établir

sur l'ADN, se scinder en demi-complexes à la réplication et reconstituer des complexes entiers sur chaque double brin fils.

Un autre aspect de la question est comment se font les changements dans le modèle d'expression d'un gène dans les cellules filles. Une possibilité souvent discutée est celle d'un changement qui pourrait être relié à la réplication de l'ADN, tirant profit de la dislocation transitoire de la structure chromatinienne. Une fois réalisée, une modification pourrait être transmise via un mécanisme de ségrégation, à moins que et jusqu'à ce qu'une autre modification intervienne.

L'ASSEMBLAGE DES NUCLÉOSOMES NÉCESSITE DES PROTÉINES NON HISTONES

Les expériences de reconstitution d'histones avec l'ADN engendrent une structure plus proche de la particule cœur que celle du nucléosome complet. A faibles concentrations, quand les particules individuelles sont reliées par un filament d'ADN, plus de 146 pb sont associées à la protéine octamère. Mais cette caractéristique n'est pas stable, puisque dès que les concentrations s'élèvent jusqu'à obtention de particules plus condensées, chaque octamère organise uniquement la longueur cœur de 146 pb. Comme si en l'absence d'une particule cœur

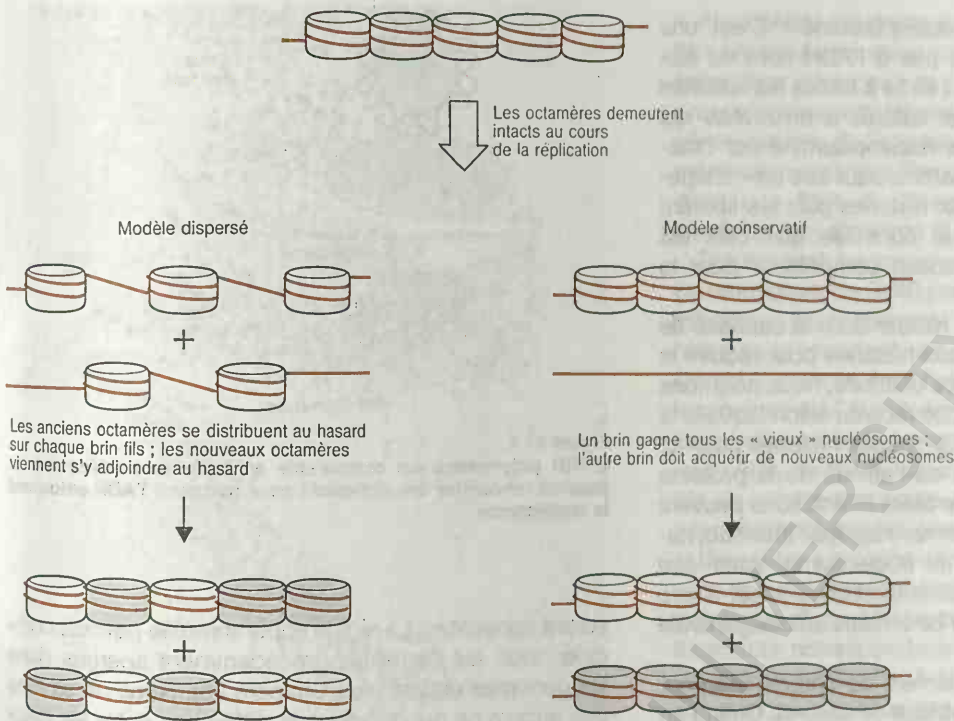


Figure 27.3

La réplication conservative des nucléosomes s'accomplirait si les « vieux » octamères se regroupaient sur le même brin fils. Mais en fait un mode dispersé est suivi dans lequel les anciens et les nouveaux octamères sont entremêlés au hasard sur les brins fils.

adjacente, un octamère mobilisait une plus grande longueur d'ADN, mais abandonnait tout ce qui n'appartient pas au cœur proprement dit dès qu'un voisin immédiat est formé.

Quel est l'élément qui module cette réaction *in vivo*, afin que la formation des nucléosomes respecte l'espace-ment caractéristique de leur génome, voire même de leur localisation particulière au sein de celui-ci ?

Un système *in vitro* qui assemble les nucléosomes sur l'ADN libre avec un espacement de 195 pb a été obtenu à partir de l'oocyte de *Xenopus*. Il s'agit d'un apport important dans la mesure où il montre que l'assemblage correct des nucléosomes peut s'effectuer *de novo* avec de l'ADN libre. L'assemblage n'est pas nécessairement associé au mécanisme de la réplication ; de plus, il est indépendant de la séquence d'ADN apportée.

Une « protéine d'assemblage » qui peut être impliquée dans la réaction, a été purifiée de l'oocyte de *Xenopus*. C'est un pentamère qui contient des sous-unités identiques de 29 000 daltons. C'est dans l'oocyte que cette

protéine est la plus abondante et elle est localisée dans le nucléoplasme. Un anticorps dirigé contre elle réagit avec les protéines des nucléoplasmes d'un grand nombre d'eucaryotes, suggérant qu'elle pourrait représenter une fonction universelle qui a été conservée au cours de l'évolution. Elle porte le nom de **nucléoplasmine**.

Dans les conditions physiologiques (faible concentration saline), en présence de nucléoplasmine, les histones peuvent s'associer à l'ADN pour engendrer des particules. Quand elles sont digérées par la nucléase micrococcale, ces particules engendrent des bandes d'ADN de 146 pb et 165 pb. Ceci laisse supposer que les particules cœur sont formées ; certaines possèdent une longueur supplémentaire d'ADN mais insuffisante pour aboutir à la construction d'un nucléosome complet. La nucléoplasmine intervient donc dans le contrôle de la réaction entre les histones et l'ADN, de telle sorte qu'elle devient productive et forme des nucléosomes au lieu d'agréats au hasard. Cependant, elle n'établit par l'espacement caractéristique des nucléosomes.

Quel est le rôle de la nucléoplasmine ? C'est une protéine acide qui ne se lie pas à l'ADN libre ou aux nucléosomes intacts mais qui se lie à toutes les histones individuelles. La réaction est saturée à un niveau qui équivaut à un pentamère de nucléoplasmine par octamère d'histone. La nucléoplasmine peut être un « chaperon moléculaire » se liant aux histones puis les libérant pour l'ADN de manière plus contrôlée que cela est possible quand l'ADN n'a aucun compétiteur pour la liaison de ces protéines.

La base de son effet peut résider dans la capacité de la nucléoplasmine à se lier aux histones pour réduire la charge positive nette. Dans ce contexte, nous pourrions également mentionner une idée récente selon laquelle la modification des groupes chargés des histones pourrait intervenir dans la régulation de l'affinité de la protéine pour l'ADN (voir plus loin). *De telles interactions peuvent permettre aux histones de former l'agrégat thermodynamiquement le plus stable (le nucléosome) sans être piégées dans des intermédiaires cinétiques (c'est-à-dire d'autres complexes résultant de la haute affinité générale des histones pour l'ADN).*

La structure globale du filament de chromatine peut être importante pour l'assemblage de celle-ci. Quand un ADN circulaire est injecté dans l'oocyte, les premiers nucléosomes s'assemblent puis le filament devient superenroulé. Ces caractéristiques peuvent être reproduites dans un système *in vitro* contenant de l'ATP, qui apporte l'énergie nécessaire à la réaction de superenroulement. Cet « assemblage actif » engendre des nucléosomes dont la périodicité est de 200 pb. La purification de ce système nous permet de définir les interactions impliquées dans la détermination de la longueur d'ADN sequestré par chaque nucléosome.

Deux caractéristiques au moins de l'assemblage du nucléosome sont extrinsèques aux constituants du nucléosome. La première intervient dans le contrôle de l'affinité des histones pour l'ADN ; c'est le rôle de la nucléoplasmine. La seconde est impliquée dans l'établissement de la longueur d'ADN contenu dans le nucléosome ; ceci peut être en rapport avec le superenroulement.

La chromatine formée par assemblage actif dans l'oocyte contient deux types de matériel, discernables par la sensibilité du superenroulement à la rupture. La majorité de la chromatine contient des superenroulements prisonniers (par interaction avec des protéines). Ils ne sont pas détendus par coupure et sont libérés uniquement quand les histones sont extraites. Cependant, une partie de la chromatine (jusqu'à 30 % environ) est sous tension. Elle présente des superenroulements qui peuvent être libérés par coupure. Cette fraction contient les séquences qui

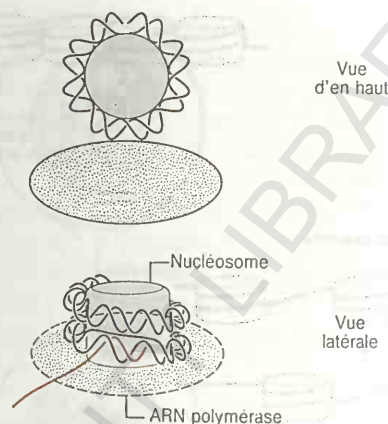


Figure 27.4

L'ARN polymérase est comparable en taille au nucléosome et pourrait rencontrer des difficultés pour parcourir l'ADN entourant le nucléosome.

seront transcrites. La région superenroulée pourrait coïncider avec les domaines précédemment aperçus dans les génomes entiers (voir Chapitre 25). Nous ne savons pas encore ce qui provoque la différence entre les deux types de matériel.

LES GÈNES TRANSCRITS SONT-ILS ORGANISÉS EN NUCLÉOSOMES ?

Lorsque nous envisageons la transcription, nous devons avoir présent à l'esprit les tailles relatives de l'ARN polymérase et du nucléosome. Les enzymes eucaryotes sont de grandes protéines, généralement supérieures à 500 000 daltons. En comparaison, celle du nucléosome est d'environ 262 000 daltons. La Figure 27.4 illustre l'approche de l'ARN polymérase vers l'ADN nucléosomal. Sans connaître cette interaction de façon détaillée, il est évident qu'elle implique le rapprochement de deux composés comparables.

Le nucléosome n'est pas un élément isolé, il est adjacent aux autres ; sans compter les deux spires d'ADN qui l'entourent. L'ARN polymérase aurait-elle un accès suffisant à l'ADN si l'acide nucléique était confiné à son enroulement habituel sur le nucléosome ? Il est difficile d'imaginer qu'au cours de la transcription la polymérase puisse suivre l'ADN autour du nucléosome.

Pendant la transcription, la polymérase qui se déplace le long de la matrice est étroitement liée à une région d'environ 50 pb comprenant un segment localement déroulé d'à peu près 12 pb. La nécessité de désenrouler

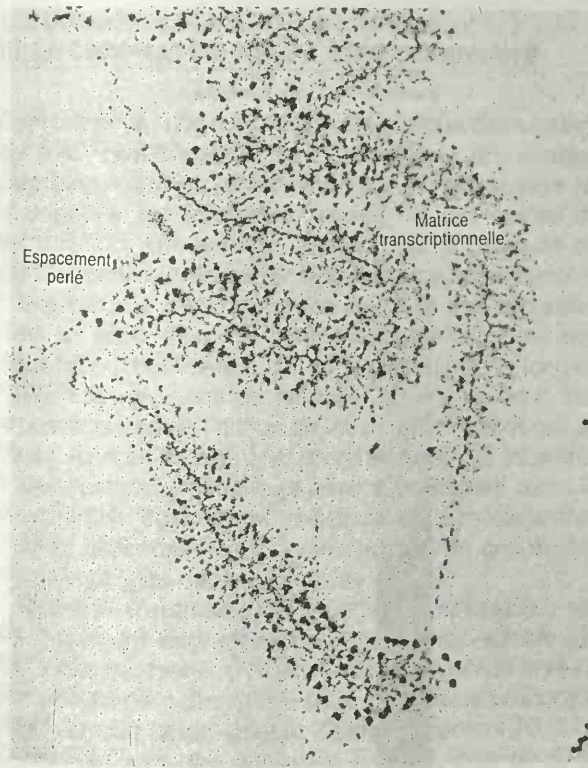


Figure 27.5
L'axe étendu d'une unité de transcription d'ADNr alterne avec des fragments intercalaires non transcrits qui ne sont que légèrement moins étendus.

Photographie aimablement fournie par Charles Laird.

l'ADN rend très improbable le fait que le segment engagé avec l'ARN puisse rester à la surface du nucléosome.

Par conséquent, il semble inévitable que la transcription induit un changement structural. Ainsi la première question qui se pose à propos de la structure des gènes actifs est de savoir si l'ADN qui est transcrit reste organisé en nucléosomes. Si les octamères d'histones sont déplacés, restent-ils d'une manière ou d'une autre liés à l'ADN transcrit ?

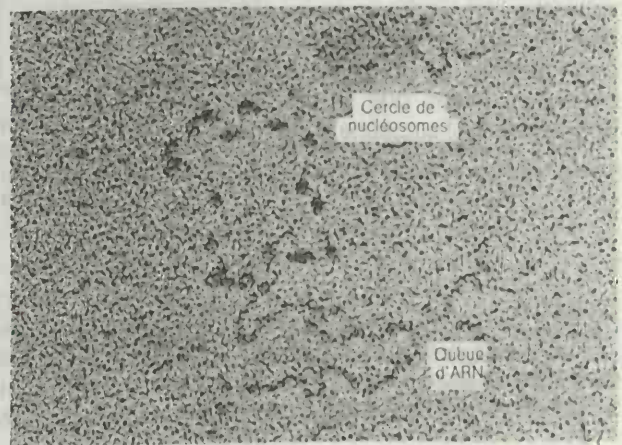
Des tentatives de visualisation de gènes en cours de transcription ont produit des résultats variés. Dans les gènes qui codent l'ARNr et qui sont intensivement transcrits, comme le montre la **Figure 27.5**, la condensation extrême des ARN polymérases rend difficile la visualisation de l'ADN. Le facteur de condensation de l'ADN peut être calculé en divisant la longueur mesurée de l'axe de la matrice de transcription par la longueur mesurée de l'axe de la

Ainsi l'ADN est presque totalement étendu ; il ne peut être organisé en nucléosomes. Les espaces non transcrits entre les matrices de transcription semblent être presque également détendus. L'état de l'ADN dans ces gènes est loin de l'organisation compacte qui serait vue même sur un simple filament de nucléosomes adjacents dans une fibre de 10 nm (qui aurait un rapport de condensation d'environ 6).

D'autre part, les complexes de transcription des minichromosomes de SV40 peuvent être extraits des cellules infectées. Ils contiennent le complément habituel des histones et montrent une structure en forme de grains. On peut, comme dans l'exemple de la **Figure 27.6**, voir les chaînes d'ARN s'étendre à partir du minichromosome. Ceci indique que la transcription peut s'effectuer tandis que l'ADN de SV40 est organisé en nucléosomes. Bien sûr, le minichromosome de SV40 est transcrit moins intensivement que les gènes de l'ARNr.

Une autre approche consiste à digérer la chromatine avec la nucléase micrococcalle, puis à utiliser une sonde spécifique d'un gène ou de gènes pour déterminer si les fragments correspondants sont présents dans l'échelle de 200 pb habituelle, à la concentration attendue. Les conclusions que nous pouvons tirer de ces expériences sont limitées mais importantes. *Les gènes qui sont en cours de transcription contiennent des nucléosomes à la même fréquence que les séquences non transcrites.* Donc, les gènes n'adoptent pas nécessairement une forme alternative d'organisation afin d'être transcrits.

Figure 27.6
Un minichromosome de SV40 peut être transcrit.
Photographie aimablement fournie par Pierre Chambon.



Mais puisque la proportion du gène associée aux ARN polymérases peut être plutôt faible, ceci ne révèle pas ce qu'il advient au niveau des sites où l'enzyme est engagée. Peut-être retiennent-ils leurs nucléosomes ; peut-être ces derniers sont-ils déplacés temporairement lorsque l'ARN polymérase passe, puis reformés immédiatement après.

Ces expériences montrent aussi que les gènes actifs sont plus rapidement digérés en fragments monomériques que les séquences inactives. Cela peut signifier que, bien qu'organisée en nucléosomes, la structure des gènes actifs est d'une certaine manière plus exposée.

Une indication que des changements interviennent dans la structure de la fibre chromatinienne durant la transcription est offerte par l'exemple de certains gènes du choc thermique de *D. melanogaster*. Ces gènes sont transcrits plutôt rarement avant un choc thermique et dans ces conditions ils montrent une échelle habituelle et semblable à celle de la digestion par la nucléase micrococcale. Quand ils sont activés par un choc thermique, ils sont transcrits intensivement. La **Figure 27.7** montre que l'activation provoque un étalement considérable de l'échelle, impliquant que l'organisation nucléosomale a été perturbée.

Nous ne connaissons pas la nature de ce changement. L'échelle peut disparaître parce qu'elle est obscurcie par la présence de l'ARN polymérase (ou d'autres protéines), parce qu'il y a rupture d'une structure d'un ordre supérieur ou parce que les nucléosomes ont été modifiés ou même déplacés tous ensemble.

Il est clair qu'un changement structural important intervient quand un gène est intensivement transcrit. Dans le cas des gènes de l'ARNr, tout se passe comme si les nucléosomes étaient déplacés. Mais il pourrait s'agir d'un cas exceptionnel. Ceci pourrait s'accorder avec la présence de nucléosomes dans les gènes moins activement transcrits en supposant que l'ARN polymérase déplace le nucléosome au point de transcription mais que l'histone octamère retrouve immédiatement sa position à moins que la présence d'une autre ARN polymérase ne l'empêche d'agir ainsi.

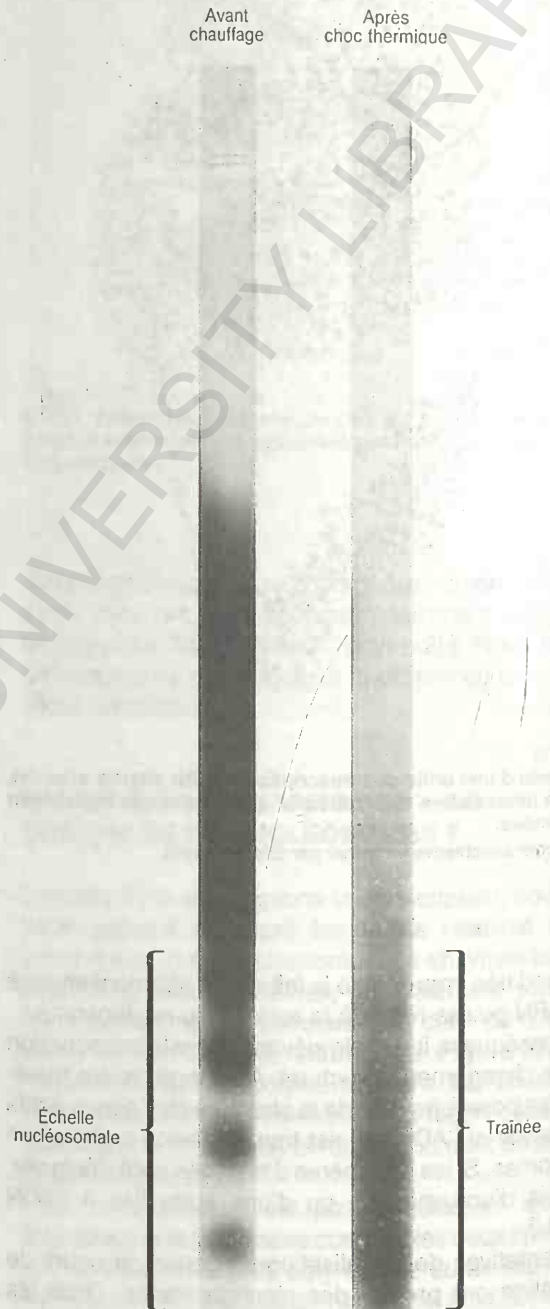


Figure 27.7

L'échelle micrococcale est évidente avant l'activation des gènes du choc thermique mais devient diffuse peu après le début d'une intense transcription.

Photographie aimablement fournie par Sarah Elgin.

LES DOMAINES SENSIBLES À L'ADNase DE LA CHROMATINE APTE À ÊTRE TRANSCRITE

Étant donné qu'une minorité caractéristique des séquences d'ADN est transcrite dans toute cellule, une certaine particularité doit au moins distinguer les promoteurs (et si possible) les longueurs totales de ces unités de transcription. Un élément doit indiquer à l'appareil de transcription qu'un locus est disponible pour l'expression.

Une certaine perturbation de structure doit apparaître dans un gène quand il est transcrit, ne serait-ce que résultant du mouvement de l'ARN polymérase le long de l'ADN. Les variations de structure pourraient donc être la conséquence de l'action de transcription plutôt que sa cause. Ainsi en estimant les propriétés des loci transcrits il nous faut déterminer lesquelles apparaissent avant la transcription comme un prérequis et lesquelles sont induites secondairement comme le résultat des événements impliqués dans la synthèse d'ARN.

Quand la chromatine est digérée par l'ADNase I, elle est finalement dégradée en matériel acido-soluble (de très petits fragments d'ADN). La progression de la réaction globale peut être suivie sous forme de la proportion d'ADN rendue acido-soluble. *Quand seulement 10 % de l'ADN total sont devenus acido-solubles, plus de 50 % de l'ADN du gène actif ont été perdus.* Ceci suggère que les gènes actifs sont préférentiellement dégradés.

Le sort des gènes individuels peut être suivi en évaluant la quantité d'ADN restante à l'aide d'une sonde spécifique. Le protocole est détaillé dans la **Figure 27.8**. Le principe en est que la disparition d'une bande particulière indique que la région correspondante d'ADN a été dégradée par l'enzyme.

La **Figure 27.9** montre ce qu'il advient des gènes de la β -globine et du gène de l'ovalbumine dans la chromatine extraite de globules rouges de poulet (dans lesquels les gènes de la globine sont exprimés et celui de l'ovalbumine inactif). Les fragments de restriction représentant les gènes de la β -globine sont rapidement perdus, tandis que ceux qui représentent le gène de l'ovalbumine ne sont que légèrement dégradés. (Le gène de l'ovalbumine est en fait digéré à la même vitesse que la masse d'ADN.)

Ainsi la masse de chromatine est relativement résistante à l'ADNase I et contient des gènes non exprimés (aussi bien que d'autres séquences). *Un gène devient relativement sensible à l'enzyme spécifiquement dans le(s) tissu(s) où il s'exprime.*

Cette susceptibilité préférentielle est-elle une caractéristique n'appartenant qu'aux gènes activement exprimés, tels que la globine, ou à tous les gènes actifs ? Des

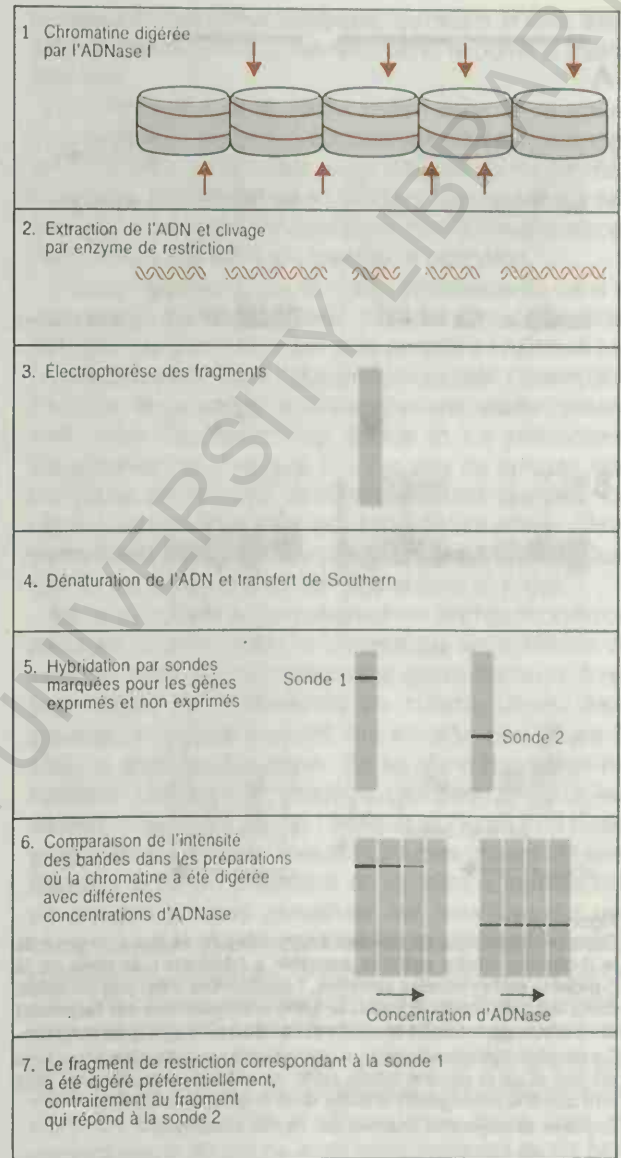


Figure 27.8

La sensibilité à l'ADNase I peut être mesurée en déterminant le degré de disparition du matériel qui s'hybride avec une sonde particulière.

expériences qui utilisent des sondes d'ADNc qui représentent la population entière d'ARNm cellulaire suggèrent que tous les gènes actifs, qu'ils codent pour des ARNm abondants ou rares, sont préférentiellement sen-

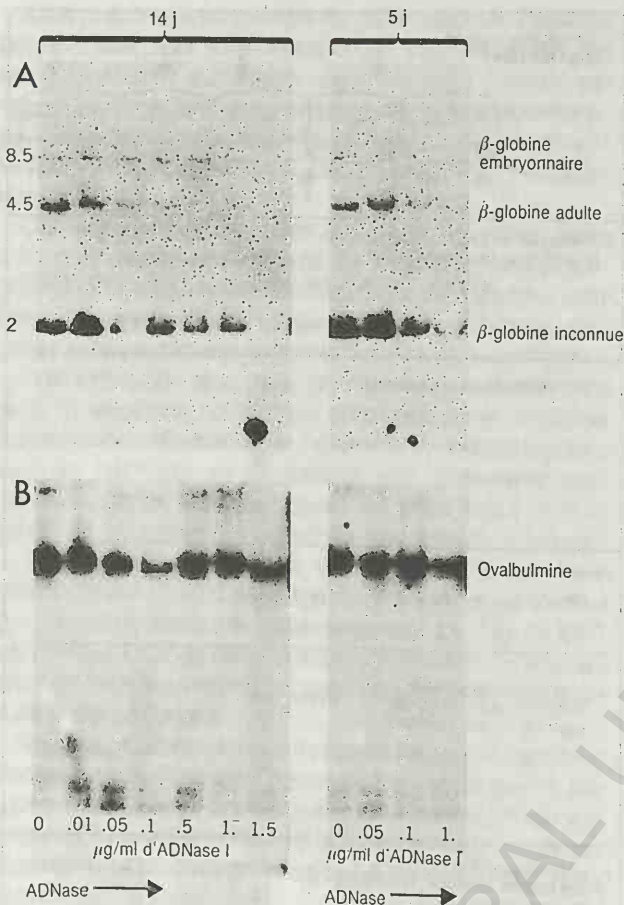


Figure 27.9

Dans la chromatine de cellules érythroïdes de 14 jours, le gène de la β -globine adulte est très sensible à l'ADNase I, le gène de la β -globine embryonnaire sensible, l'ovalbumine n'est pas sensible. Dans les cellules de 5 jours, le gène embryonnaire est fortement sensible, le gène adulte sensible et l'ovalbumine n'est pas sensible. (La sensibilité intermédiaire de la globine embryonnaire dans les cellules adultes et de la globine adulte dans les cellules embryonnaires peut être due à la propagation d'effets dans le groupe de gènes.)

Données aimablement fournies par Harold Weintraub.

sibles à l'ADNase I. (Il peut cependant exister des différences dans les degrés de sensibilité.) Puisque les gènes qui s'expriment rarement n'ont probablement que peu de molécules d'ARN polymérase réellement engagées dans la transcription à un moment donné, ceci implique que la sensibilité à l'ADNase I ne résulte pas de la transcription en elle-même, mais est une caractéristique des gènes qui sont capables d'être transcrits.

Quelle est l'étendue de la région préférentiellement sensible ? Elle peut être déterminée en utilisant des

sondes qui représentent les régions bordantes aussi bien que l'unité de transcription elle-même. La région sensible s'étend toujours sur la totalité de la région transcrite ; de chaque côté, une région supplémentaire de plusieurs kb peut montrer un niveau de sensibilité intermédiaire entre celui de la région transcrite et celui de la masse de chromatine.

La sensibilité à l'ADNase détermine un **domaine** chromosomique, région de structure altérée comprenant au moins une unité de transcription active qui s'étend peut-être au-delà. Il est à noter que l'utilisation du terme « domaine » n'implique pas nécessairement une connexion avec les domaines structuraux déterminés par les boucles de la chromatine ou des chromosomes.

Implicitement, le concept critique à propos de la description de ce domaine est qu'une région fortement sensible à l'ADNase I s'étend sur une distance considérable. Nous envisageons souvent la régulation comme un mécanisme n'intervenant qu'en un site discret dans l'ADN, par exemple, dans la capacité à initier la transcription à partir du promoteur. Même si cela est exact, une telle régulation doit déterminer, ou doit être accompagnée par un changement de structure plus largement étendu. Ce pourrait être une différence entre les eucaryotes et les procaryotes.

LES HISTONES SONT MODIFIÉES TRANSITOIREMENT

Toutes les histones sont modifiées en liant de façon covalente des parties supplémentaires aux groupes libres de certains acides aminés. Le groupe aminé libre (ϵ) de la lysine est acétylé et méthylé. Comme l'indique la Figure 27.10, cela fait disparaître la charge positive portée par la forme NH_3^+ du groupe. La méthylation intervient également sur l'arginine et l'histidine. Le groupe hydroxyle de la sérine et aussi de l'histidine est phosphorylé, le groupe phosphate apportant alors une charge négative (voir Figure 27.10).

Toutes ces modifications affectent les résidus internes et sont transitoires. Elles apparaissent à un moment du cycle cellulaire et disparaissent (habituellement) à un autre moment. Du fait qu'elles provoquent un changement de charge de la molécule protéique, elles ont été considérées comme potentiellement capables de changer les propriétés fonctionnelles des histones. Actuellement, aucun argument ne permet de relier ces changements aux fonctions de la chromatine bien que certaines corrélations soient tout à fait évidentes.

Dans les cellules synchronisées en culture, les histones

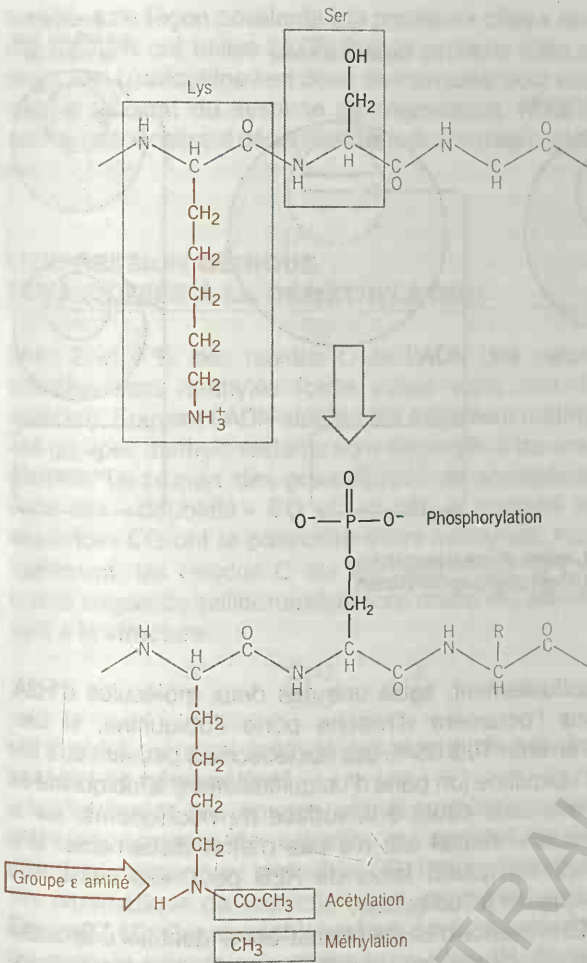


Figure 27.10
La charge positive globale d'une protéine est réduite par l'acétylation de la lysine ou la phosphorylation de la sérine.

du cœur préexistantes aussi bien que celles nouvellement synthétisées semblent être acétylées et méthylées au cours de la phase S (quand l'ADN est répliqué et les histones également synthétisées). Durant ce cycle cellulaire, les groupes modifiés sont libérés ultérieurement. Ces événements se répètent à chaque cycle cellulaire et leur chronologie relative est résumée dans la Figure 27.11.

La coïncidence entre modification et réplication suggère que l'acétylation (ou la méthylation) pourrait être en rapport avec l'assemblage des nucléosomes. Selon une hypothèse, la réduction des charges positives des histones pourrait abaisser leur affinité pour l'ADN, permettant un meilleur contrôle de la réaction. Mais l'observation de

la reconstitution de nucléosomes, au moins *in vitro*, avec des histones non modifiées rend cette hypothèse moins probable.

Un obstacle à cette étude réside dans le caractère transitoire de l'acétylation. La difficulté peut être surmontée par addition d'acide butyrique aux cellules croissantes en culture. Ce traitement, en inhibant la déacétylase des histones, provoque l'accumulation des nucléosomes acétylés. Toutes les histones cœur sont acétylées.

L'acétylation est associée à des changements dans la chromatine similaires à ceux trouvés dans l'activation génique. La chromatine est plus sensible à l'ADNase I et (vraisemblablement) à la nucléase micrococcale. Cependant, il n'a pas été possible de démontrer une relation essentielle entre l'activation des gènes et ce phénomène d'acétylation, et d'ailleurs, il n'y a pas de preuves que les gènes actifs soient préférentiellement acétylés. Ce résultat nous indique donc que l'acétylation affecte néanmoins la structure de la chromatine mais que la signification de ce changement n'est pas encore élucidée.

Un cycle de phosphorylation et de déphosphorylation intervient avec H1, mais sa chronologie est différente de celle du cycle de modification des autres histones. Avec des cellules de mammifères en culture, un ou deux groupes phosphate peuvent être introduits en phase S. Mais la plus grande partie de la phosphorylation est l'addition ultérieure de plusieurs groupes qui porte leur nombre total à 6. Comme l'indique la Figure 27.11 ceci se produit à la mitose. Tous les groupes phosphate sont libérés à la fin du processus de division. L'introduction de certains groupes phosphate est catalysée par une phosphokinase dont l'activité s'accroît intensément à chaque début de mitose. On connaît encore peu de chose sur la phosphatase qui libère ces groupes ultérieurement.

La chronologie de la phosphorylation majeure d'H1 a laissé rapidement envisager qu'elle pourrait être impliquée dans la condensation mitotique. Ce qui est certainement en accord avec la dépendance de la fibre de chromatine de 30 nm vis-à-vis de la présence de H1 (voir Chapitre 26).

Par contraste avec les modifications qui ne font intervenir l'addition que de simples groupes, un polypeptide entier peut être ajouté à H2A. La caractérisation d'une protéine présente à l'origine dans la fraction non histone de la chromatine de foie de rat a permis de mettre en évidence cette modification extraordinaire. Cette protéine possède la même séquence d'acides aminés en C-terminal que H2A, mais elle a deux séquences N-terminal comme l'illustre la Figure 27.12. Une séquence est celle de H2A. L'autre est celle de la protéine ubiquitine, dont l'appellation reflète la présence ubiquitaire dans les cellu-

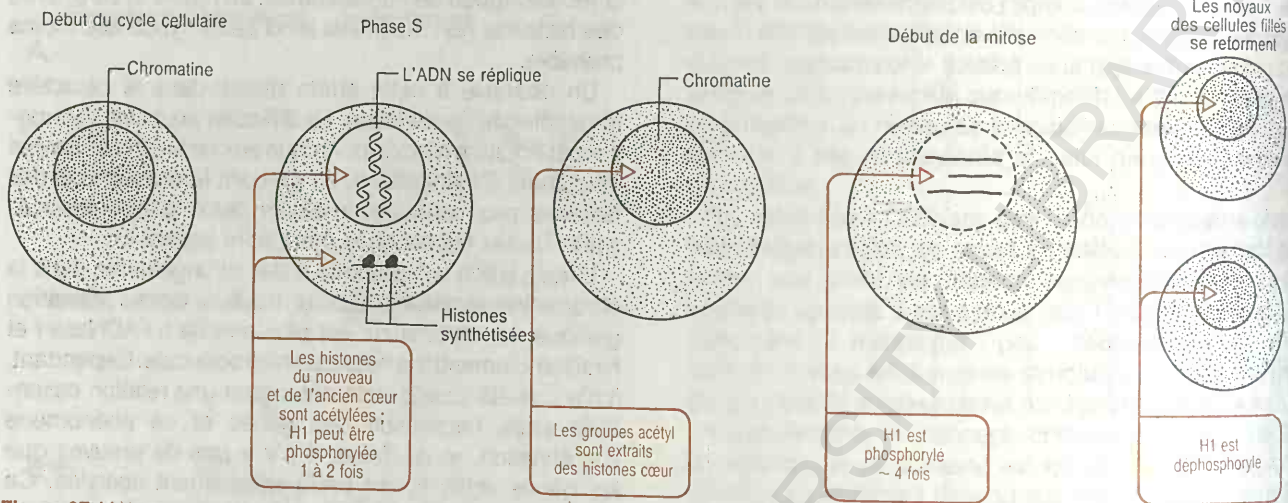


Figure 27.11

L'acétylation de l'histone cœur intervient au cours de la synthèse de l'ADN, suivie d'une désacétylation ; la phosphorylation majeure d'H1 réalisée en début de mitose est suivie d'une déphosphorylation à la fin de la mitose.

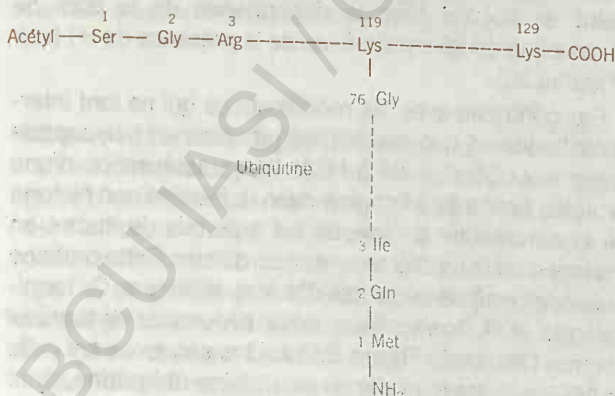
les de bactéries jusqu'à des cellules de mammifères. H2A-ubiquitine est appelée **UH2A**.

Une liaison isopeptide est formée entre le C-terminal d'une glycine dans l'ubiquitine et le ϵ -NH₂ libre de la lysine en position 119 de H2A. (Liaison *isopeptide* indique que le groupe NH₂ n'est pas celui qui intervient habituellement dans la formation d'une liaison peptidique.) L'ubiquitine est une protéine acide dont le contenu en acides glutamique et aspartique réduit dans la protéine conjuguée le rapport base/acide.

5 à 15 % de H2A peuvent être sous la forme UH2A.

Figure 27.12

UH2A est constituée d'ubiquitine liée à H2A.



Habituellement, seule une des deux molécules d'H2A, dans l'octamère d'histone, porte l'ubiquitine, si bien qu'environ 10 à 35 % des nucléosomes peuvent être liés à l'ubiquitine (on parle d'ubiquitination). L'ubiquitine se trouve sans doute à la surface du nucléosome, sur la structure duquel elle n'a pas d'effet discernable. Une proportion plutôt faible de H2B peut, elle aussi, être conjuguée à l'ubiquitine.

La présence d'UH2A permet-elle d'identifier une classe spéciale de nucléosomes ? Environ 20 % des nucléosomes de la chromatine de *D. melanogaster* possèdent UH2A ; et leurs séquences d'ADN peuvent être étudiées par transfert de fractions nucléosomales et par hybridation avec des sondes spécifiques. Avec des sondes qui représentent des gènes transcrits, 50 % du matériel se trouvent dans la fraction ubiquitine. Avec l'ADN satellite, on ne retrouve que moins de 4 % de matériel ubiquitine. Donc UH2A tend à être concentrée dans les nucléosomes des séquences transcrites et semble être exclue de l'hétérochromatine.

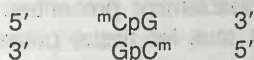
Le rôle de l'ubiquitine est encore mal connu. A la mitose, elle semble être libérée de la chromatine, phénomène déconcertant puisqu'il soulève le problème de la perpétuation à travers le cycle cellulaire des régions particulières chargées d'ubiquitine.

Un rôle possible de l'ubiquitine a été signalé dans le cytoplasme où elle fait partie d'un système de dégradation protéique. Une ou plusieurs parties de l'ubiquitine

sont liées de façon covalente à la protéine « cible » dans une réaction qui utilise l'ATP. Puis la protéine cible est dégradée. L'ubiquitine sert donc de marqueur pour identifier le substrat du système de dégradation. Nous ne savons pas si cela est dépendant de mécanismes nucléaires.

L'EXPRESSION GÉNIQUE EST ASSOCIÉE À LA DÉMÉTHYLATION

Entre 2 et 7 % des résidus C de l'ADN des cellules animales sont méthylés (cette valeur varie avec les espèces). Souvent l'ADN satellite est largement méthylé. Les groupes méthyle restants sont dispersés à travers le génome. La plupart des groupes méthyle apparaissent dans des « doublets » CG et, en fait, la majorité des séquences CG ont la possibilité d'être méthylées. Habituellement, les résidus C sur les deux brins de cette courte séquence palindromique sont méthylés, aboutissant à la structure :



Un doublet qui n'est méthylé que sur un de ses deux brins est dit **hémiméthylé**.

La distribution des groupes méthyle peut être examinée grâce aux enzymes de restriction qui coupent les sites cibles qui contiennent le doublet CG. Deux types d'activité enzymatique de restriction sont comparés dans la **Figure 27.13**. Ces **isoschisomères** sont des enzymes qui coupent dans la même séquence cible d'ADN mais

qui répondent différemment selon l'état de méthylation de cette séquence.

Dans ce double exemple, l'enzyme HpaII coupe la séquence CCGG (la séquence n'est écrite que pour un brin d'ADN). Mais si le second C est méthylé, l'enzyme ne peut plus reconnaître le site. En revanche, l'enzyme MspI coupe le même site cible quelque soit l'état de méthylation de ce C. MspI peut donc être utilisée pour identifier toutes les séquences CCGG et HpaII détermine uniquement si elles sont ou non méthylées.

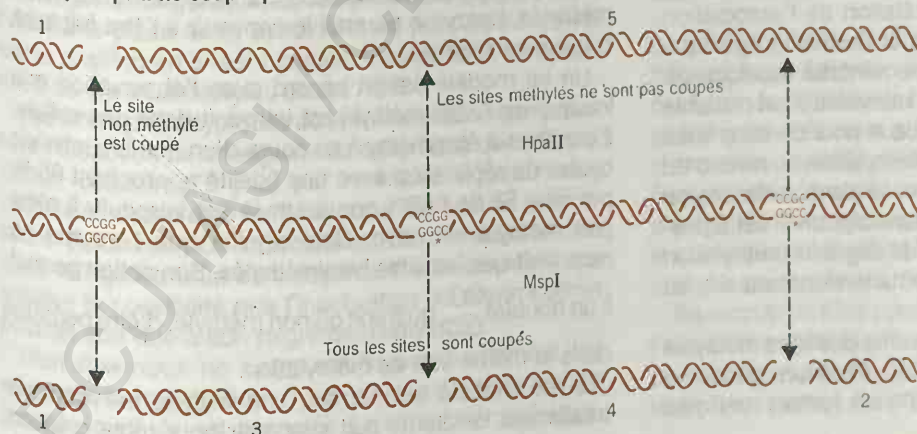
Si le substrat est un ADN non méthylé, les deux enzymes engendreront les mêmes bandes de restriction. Dans le cas contraire, les positions modifiées ne seront pas coupées par HpaII. Pour chaque position dans ce cas, un fragment HpaII plus long remplace deux fragments MspI. Un exemple est donné dans la **Figure 27.14**.

L'utilisation des enzymes de restriction permet d'identifier des sites qui ne contiennent que quelques-unes des séquences méthylées, mais il est sûr que leur comportement est typique de l'ensemble du génome. Le modèle de méthylation est-il invariant ou est-il soumis à des altérations selon les circonstances ?

Des gènes variés ont un comportement similaire. Certains de leurs sites sont méthylés dans tous les tissus examinés alors que d'autres sites ne sont méthylés dans aucun tissu. *Une minorité de sites sont méthylés dans les tissus dans lesquels le gène ne s'exprime pas, alors qu'ils ne sont pas méthylés dans les tissus où le gène est actif.* Un gène actif peut donc être décrit comme étant « sous-méthylé ».

De la même manière que nous avons étudié l'état de

Figure 27.13
L'enzyme de restriction Msp I coupe toutes les séquences CCGG qu'elles soient méthylées ou non, alors que Hpa II ne coupe que les tétramères CCGG non méthylés.



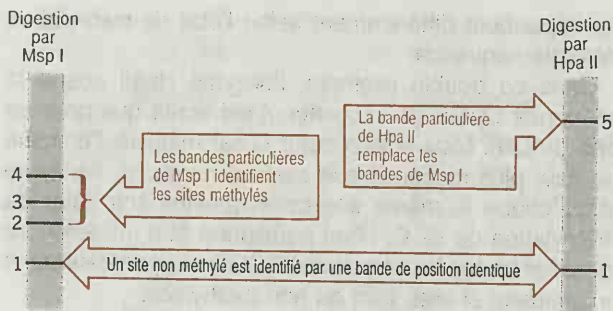


Figure 27.14

Les résultats des coupures par Msp I et Hpa II sont comparés par électrophorèse sur gel des fragments.

La numérotation des bandes se réfère aux fragments identifiés dans la Figure 27.13.

méthylation des séquences endogènes, nous pouvons comparer les résultats obtenus après introduction d'ADN méthylé ou non méthylé dans de nouvelles cellules hôtes. De telles expériences ont permis d'établir une corrélation évidente : le gène méthylé est inactif mais le gène non méthylé est actif.

L'étendue de la région hypométhylée *in vivo* a été déterminée dans le groupe de gènes de l' α globine de poulet. Dans les cellules érythroïdes adultes, l'hypométhylation est restreinte aux sites qui s'étendent, à peu près, de 500 pb en amont du premier des deux gènes adultes à environ 500 pb en aval du second. Les sites d'hypométhylation sont présents sur la totalité de la région y compris l'intervalle entre les gènes. La région d'hypométhylation coïncide plutôt bien avec la région la plus sensible à l'ADNase I. Ceci suggère que la sous-méthylation peut être caractéristique d'un domaine qui contient un ou plusieurs gènes transcrits.

Notre problème dans l'interprétation de l'association générale entre l'hypométhylation et l'activation génique réside dans le fait que seule une minorité (quelquefois très faible) de ces sites méthylés intervient. Il est possible que l'état de méthylation soit critique pour certains sites spécifiques ; par exemple, la déméthylation au niveau du promoteur pourrait intervenir en rendant cette région disponible pour l'initiation de la transcription. Il est également possible qu'une diminution du degré de méthylation fasse partie d'un changement structural nécessaire au mécanisme de la transcription.

Dans certains gènes il existe en effet des sites méthylés en 5' qui sont reliés à l'expression. Plusieurs gènes ne sont pas méthylés à l'extrémité 5' mais restent méthylés

à l'extrémité 3' quand ils s'expriment ; comme pour les autres changements dans la chromatine, il semble probable que l'absence de groupes méthyle soit associée à la capacité à être transcrit plutôt qu'au mécanisme de transcription en lui-même. L'importance de la déméthylation du promoteur devrait être relativement facile à tester directement, puisqu'il devrait être possible de construire *in vitro* des gènes qui diffèrent uniquement par leur degré de méthylation en 5'.

Il existe des exceptions à la relation générale que nous avons décrite. Certains gènes peuvent s'exprimer même s'ils sont très largement méthylés. Toute relation entre méthylation et expression n'est donc pas universelle.

Une grande attention a été portée au mécanisme du maintien ou du changement de l'état de méthylation. Les gènes sensibles à un contrôle spécifique de tissu sont inactifs dans le sperme ; ils sont méthylés à la fois à des sites constants (modifiés dans tous les tissus) et à des sites variables (ceux non méthylés spécifiquement dans les tissus où ils s'expriment). Donc, l'absence de certains groupes méthyle dans l'état actif représente la perte de modifications probablement présentes. Les gènes qui s'expriment dans tous les tissus peuvent être à l'état déméthylé dans le sperme, l'absence de groupes méthyle pourrait fournir un signal pour leur expression.

A partir de quel moment, quand ils cessent de s'exprimer les gènes cellulaires regagnent-ils des groupes méthyles ? Nous aimerions savoir comment s'effectue la sélection des séquences cibles qui interviennent dans les différents états de méthylation spécifiques de tissu.

Une représentation simple de la perpétuation des sites méthylés consiste à supposer que l'ADN méthylase agit sur l'ADN hémiméthylé. Comme le montre la Figure 27.15, la réplication d'un doublet CG complètement méthylé aboutit à la formation de deux doubles brins fils hémiméthylés. En reconnaissant chacun des sites hémiméthylés, l'enzyme pourrait le convertir à l'état habituel, totalement méthylé.

Un tel modèle est en accord avec l'observation que lorsque de l'ADN méthylé est introduit dans une cellule, il continue à être méthylé au cours d'un nombre infini de cycles de réplication avec une fidélité approchant 95 % par site. Si de l'ADN non méthylé est introduit, il n'est pas méthylé *de novo*. Ceci signifie que l'enzyme ne reconnaît que les sites hémiméthylés. Son action permet

à un doublet $^{CG}_{CG}$, qu'il soit ou non méthylé, d'être perpétué dans le même état de méthylation.

Si ce modèle est correct, une activité enzymatique totalement différente doit intervenir pour toute création

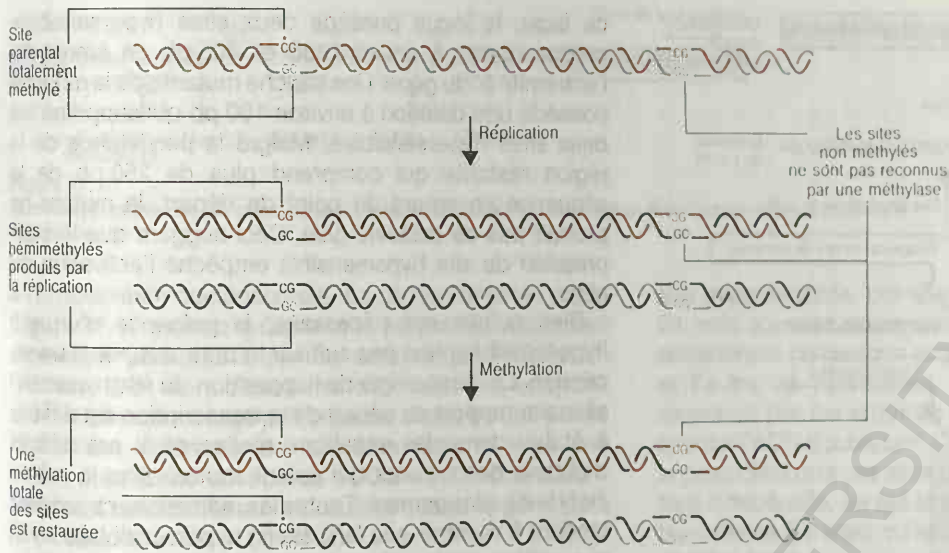


Figure 27.15

L'état des sites méthylés pourrait être perpétué par une enzyme qui ne reconnaît comme substrat que les sites hémiméthylés.

de nouveaux sites de méthylation (par exemple, si l'état de méthylation d'un gène est restauré quand la transcription cesse).

La condition méthylée peut être perdue de deux façons. Les groupes méthyle pourraient être supprimés activement par une déméthylase. Ou bien la méthylation pourrait simplement négliger d'intervenir sur un site hémiméthylé engendré par la réplication. L'un des doubles brins d'ADN produit par la réplication qui suit serait alors dénué de site méthylé.

Des expériences avec la 5-azacytidine mettent en évidence, indirectement, que la déméthylation peut aboutir à l'expression génique. Le composé est incorporé dans l'ADN à la place de la cytidine et ne peut être méthylé car la position en 5' est bloquée. Ceci conduit à l'apparition de sites déméthylés dans l'ADN.

Les effets phénotypiques de la 5-azacytidine comprennent l'induction de changements dans l'état de différenciation cellulaire ; il y a, par exemple, induction des cellules musculaires à se développer à partir de précurseurs cellulaires non musculaires. Le composé active aussi les gènes d'un chromosome X silencieux, ce qui soulève la possibilité que l'inactivation du chromosome pourrait être due à son degré de méthylation.

Bien que nous ne comprenions pas totalement les effets de la méthylation, il est clair que l'absence de groupes méthyle est associée à l'expression génique.

Cependant, il est difficile de supposer que l'état de méthylation fournisse le moyen général de contrôle de l'expression génique. Dans le cas de *D. melanogaster* (et d'autres insectes Diptères), il n'existe pas de méthylation de l'ADN. Les autres différences entre la chromatine inactive et la chromatine active semblent être les mêmes que dans les espèces qui montrent une méthylation. Ainsi, chez la drosophile, la méthylation est soit superflue, soit remplacée par un autre mécanisme.

LES SITES HYPERSENSIBLES À L'ADNase MODIFIENT LA STRUCTURE DE LA CHROMATINE

Le premier effet de la digestion de la chromatine par une très faible concentration d'ADNase est l'introduction de cassures dans le double brin à des **sites hypersensibles** spécifiques. Les sites sont créés par la structure de la chromatine (spécifique du tissu). Leur localisation peut être déterminée par la technique de marquage indirecte d'extrémité dont nous avons parlé précédemment dans le contexte de l'étude de la mise en phase des nucléosomes. Le protocole est récapitulé dans la Figure 27.16.

Beaucoup de sites hypersensibles sont en rapport avec l'expression génique. Tous les gènes actifs ont un site de coupure ou quelquefois plus dans la région immédiatement en amont du promoteur. Le site hypersensible

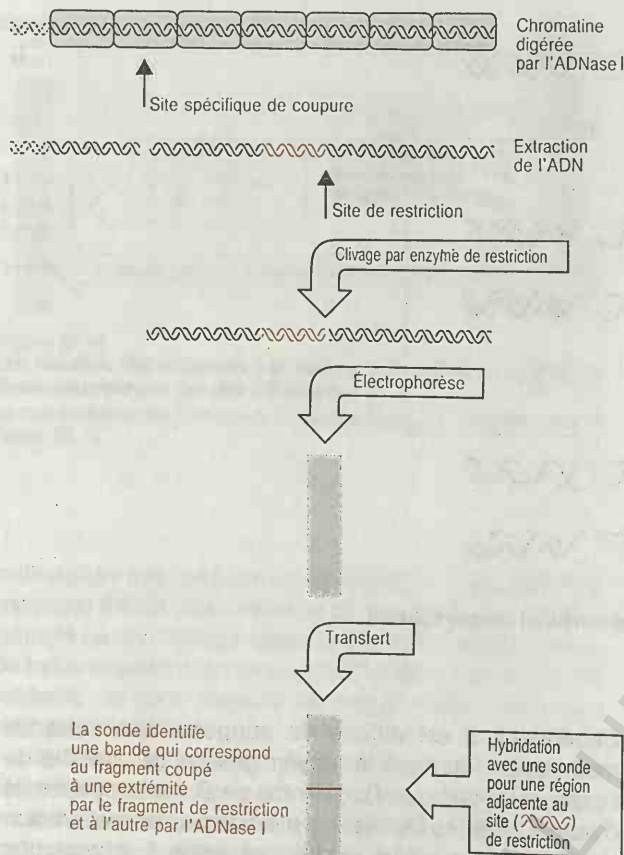


Figure 27.16

Le marquage indirect d'extrémités détermine la distance entre un site hypersensible à l'ADNase I et un site de clivage de restriction. La chromatine est digérée par l'ADNase I, l'ADN est isolé puis clivé par une enzyme de restriction. Le matériel produit par double digestion est soumis à une électrophorèse puis hybridé avec une sonde qui représente une région adjacente au site de restriction. L'existence d'un site de coupure particulier pour l'ADNase engendre un fragment distinct identifié par une bande dont la taille indique la distance entre le site hypersensible à l'ADNase I et le site de restriction.

n'est trouvé que dans la chromatine des cellules dans lesquelles le gène est en train de s'exprimer ; il n'apparaît pas quand le gène est inactif. Par exemple, dans les gènes de la globine, les sites hypersensibles sont décelés en amont des gènes embryonnaires dans les cellules embryonnaires et non dans les cellules adultes et vice-versa*.

La corrélation entre la transcription et les sites hypersensibles en amont est fortement consolidée par les données sur le gène *Sgs 4* de *D. melanogaster*. Ce gène code pour la protéine glue des glandes salivaires. Dans

ce tissu, le locus possède deux sites hypersensibles prédominants, à environ 330 et 405 pb en amont de l'extrémité 5' du gène. Une souche mutante de la mouche possède une délétion à environ 100 pb qui supprime les deux sites hypersensibles. Malgré la persistance de la région restante qui comprend plus de 250 pb de la séquence en amont du point de départ, le mutant ne produit pas de protéine glue. Ceci suggère que la suppression du site hypersensible empêche l'activation du gène.

Bien qu'elle soit nécessaire, la présence d'un site hypersensible n'est pas suffisante pour assurer la transcription. La chronologie de l'apparition du site hypersensible par rapport au départ de la transcription est difficile à établir dans des situations réelles où il est difficile d'obtenir des cellules qui soient toutes dans le même état de développement. Toutes les indications cependant tendent à montrer que le(s) site(s) hypersensible(s) en 5' apparaît(aissent) avant que la transcription ne commence, très probablement comme un préalable à l'initiation.

La stabilité des sites hypersensibles est mise en évidence par les propriétés des fibroblastes de poulet transformés par des virus oncogènes sensibles à la température. Ces expériences tirent profit d'une propriété inhabituelle : bien que les fibroblastes n'appartiennent pas au lignage érythroïde, la transformation des cellules à température normale induit l'activation des gènes de la globine. Les gènes activés ont des sites hypersensibles. Si la transformation est effectuée à une température plus élevée (non permissive), les gènes de la globine ne sont pas activés et les sites hypersensibles n'apparaissent pas. Quand les gènes de la globine ont été activés par transformation à basse température, ils peuvent être inactivés par élévation de la température. Mais les sites hypersensibles persistent au moins durant les 20 divisions cellulaires suivantes.

Ce résultat démontre à nouveau que l'acquisition d'un site hypersensible n'est qu'une des caractéristiques nécessaires à l'initiation de la transcription ; et il implique que les mécanismes qui interviennent dans l'établissement d'un site hypersensible sont distincts de ceux qui servent à le perpétuer. Une fois le site établi, il est perpétué à travers la réplication en l'absence de circonstances nécessaires à l'induction. La disparition d'un site hypersensible nécessiterait-elle une intervention spécifique ?

Un site hypersensible identifie une région inférieure à 200 pb dont l'accès n'est pas limité selon le mode typique

* N. d. T. : Il existe en fait plusieurs classes de sites hypersensibles : les uns sont en effet associés à des gènes activement transcrits, d'autres sont constamment présents dans l'ADN de tissus capables de transcrire le gène, que ce dernier soit ou non transcrit ; l'absence de transcription du gène peut être due, par exemple, à l'état hormonal. En d'autres termes, certains sites reflètent l'aptitude du gène à être transcrit dans un contexte donné de différenciation cellulaire, alors que d'autres sont étroitement associés à la transcription réelle du gène.

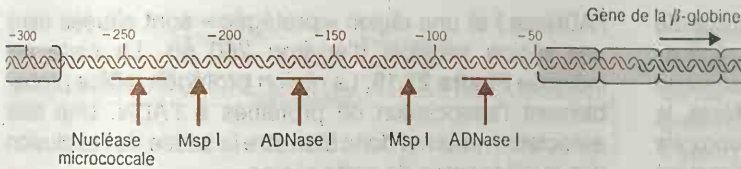


Figure 27.17

Le site hypersensible du gène de la β -globine de poulet est une région étendue sensible à plusieurs nucléases.

à l'organisation nucléosomale. La carte présentée par la Figure 27.17 montre qu'un tel site est préférentiellement digéré par plusieurs enzymes, comprenant l'ADNase I, l'ADNase II et la nucléase micrococcale. Les enzymes ont des sites de clivage préférentiels situés en des points légèrement différents de la même région. Ainsi, une région qui s'étend environ de -70 à -270 est préférentiellement accessible aux nucléases quand le gène peut être transcrit.

Comment les sites hypersensibles sont-ils établis ? L'analyse de leur structure nécessite un système *in vitro*. La région hypersensible du gène de la β -globine de poulet adulte a été reconstruite dans un plasmide. Recombinée aux histones, en présence d'extraits de noyaux de globules rouges, la région en question devient hypersensible. Il devrait être possible de purifier le facteur actif en utilisant comme test la capacité à engendrer une région hypersensible.

Quelle est la structure du site hypersensible ? Son accessibilité préférentielle aux nucléases indique qu'il n'est pas protégé par les octamères d'histones, ce qui n'implique pas nécessairement qu'il soit dépourvu de protéine. Une région d'ADN libre pourrait donc être vulnérable à la dégradation ; et en tout cas, pourquoi les nucléosomes en seraient-ils exclus ? Nous nous attendons à la présence d'autres protéines dans cette région. En effet, en créant un site hypersensible, un (ou des) facteur contenu dans l'extrait protège les séquences dans la région accessible, qui sont ainsi de ce fait moins accessibles *in vivo*.

S'il est ajouté après les histones, l'extrait ne peut conférer une hypersensibilité, ce qui suggère qu'il doit reconnaître directement l'ADN et d'une certaine façon reconnaître l'organisation de la région alors que les nucléosomes sont absents. Dans ces conditions, tout au moins, le constituant ne peut déplacer les nucléosomes après leur formation, ce qui nous ramène à la question du mécanisme de changement de la structure de la chromatine quand le gène doit être activé.

Les sites hypersensibles peuvent être coupés par la nucléase S1, une enzyme dont le substrat est l'ADN simple brin. Cette susceptibilité laisse supposer que l'ADN double brin en lui-même peut être modifié dans un

site hypersensible. Les sites de coupure de la nucléase S1 sont souvent situés dans des régions où un brin est entièrement ou presque entièrement constitué de purines et l'autre de pyrimidines. Les sites de coupure sont reconnus par les enzymes de restriction qui n'agissent que sur l'ADN double brin. Ces caractéristiques soulèvent la possibilité que ces sites possèdent une certaine structure inhabituelle, ce qui leur permet d'être reconnus par la nucléase S1 alors qu'ils ne forment pas réellement de simples brins.

Un site hypersensible représente un changement de structure dans la chromatine. Les séquences hypersensibles ne prennent probablement pas la structure nucléosomale habituelle, sans doute du fait que d'autres protéines sont complexées à l'ADN et empêchent la liaison des histones. Nous pouvons imaginer que ces protéines sont peut-être des facteurs de régulation.

Un préalable à la formation des sites peut être l'absence de groupes méthyle aux sites compris dans la région hypersensible. Une particularité de structure impliquée dans la formation des sites hypersensibles pourrait être le degré de superenroulement dans les régions voisines, de telle sorte que des activités topo-isomérase pourraient avoir une influence importante.

Les sites hypersensibles pourraient intervenir directement dans la fonction du promoteur, en fournissant très probablement des sites de liaison reconnus directement par l'ARN polymérase, par des facteurs de transcription ou peut-être par d'autres protéines impliquées dans le contrôle de la transcription. Alternativement, l'action de ces sites pourrait être plus générale dans l'activation génique, par exemple, le déclenchement d'un changement dans la structure locale nécessaire à la liaison de l'ARN polymérase ou même à l'activité transcriptionnelle.

Une caractéristique significative des sites hypersensibles proches des promoteurs est leur localisation à des distances variées du point de départ de la transcription. Dans le gène *Sgs* de *D. melanogaster*, le site est à plus de 250 pb en amont ; dans le gène de la β -globine de poulet, il est juste en amont et dans le gène de la thymidine kinase, il comprend toutes les séquences nécessaires à la fonction du promoteur.

Une région sensible à la nucléase et particulièrement

bien caractérisée est située sur le minichromosome de SV 40. Un court segment, proche de l'origine de réplication, juste en amont du promoteur de l'unité de transcription tardive est préférentiellement clivé par l'ADNase, la nucléase micrococcale et d'autres nucléases (y compris les enzymes de restriction). Ce fragment particulièrement sensible à la digestion a une longueur d'environ 400 pb et le segment peut être libéré sous forme d'un fragment d'ADN libre.

La microscopie électronique permet de visualiser l'état du minichromosome de SV40. Dans jusqu'à 20 % des échantillons, une interruption est visible dans l'organisation nucléosomale, comme le met en évidence la **Figure 27.18**. L'interruption est une région d'environ 120 nm de long (jusqu'à 350 pb), entourée par des nucléosomes qui occupent le reste du génome.

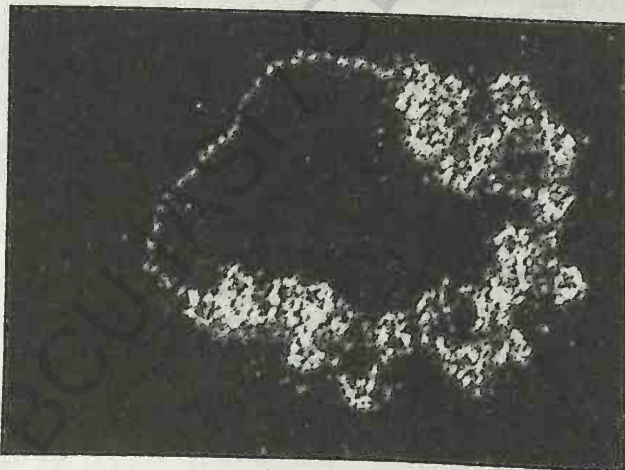
La localisation de l'interruption peut être déterminée en coupant le minichromosome circulaire avec une enzyme de restriction n'ayant qu'une seule cible connue. L'interruption visible correspond avec la région sensible à la nucléase. Ceci montre directement que la sensibilité accrue aux nucléases est associée à l'exclusion des nucléosomes. Cela ne signifie pas, qu'*in vivo*, l'ADN soit libre, car d'autres structures de nature protéique ont pu être perdues au cours de l'isolement du minichromosome.

Une interruption également sensible à la nucléase est trouvée dans le virus de polyome, et dans ce cas il a été montré que toute la région n'est pas uniformément sensible aux nucléases. Deux sites hypersensibles à

Figure 27.18

Le minichromosome de SV40 peut avoir une région dépourvue de nucléosomes.

Photographie aimablement fournie par Moshe Yaniv.



l'ADNase I et une région « protégée » sont situées dans une région sensible d'environ 260 pb. La carte est indiquée **Figure 27.19**. La région protégée reflète probablement l'association de protéines à l'ADN. Une telle association pourrait donc bien être la cause de l'exclusion des nucléosomes de cette région.

La région de SV40 ou de polyome entourant l'interruption sensible à la nucléase possède plusieurs fonctions potentielles qui comprennent l'initiation de la réplication et de la transcription ; ainsi, le rôle de l'interruption ne peut pas être simplement réduit à une activité particulière. L'interruption peut être associée avec les éléments « enhancer » qui sont situés dans cette région et qui sont nécessaires à l'action du promoteur (voir Chapitre 9).

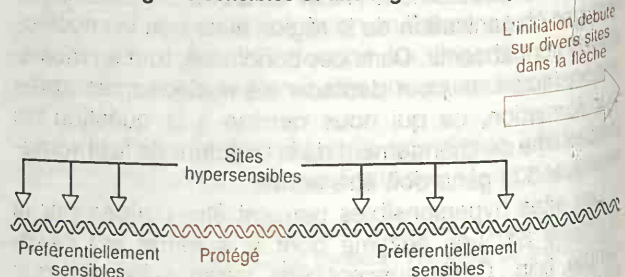
Une structure intéressante a été trouvée dans l'ADNr extrachromosomique du macro-noyau de *Tetrahymena pyriformis*. Souvenons-nous que la molécule d'ADNr est un palindrome linéaire, dans lequel les deux unités de transcription de l'ARNr sont séparées et transcrites dans des directions opposées. L'organisation de l'ADNr est représentée dans la **Figure 27.20**. Deux origines de réplication, courtes séquences où la réplication de l'ADN est initiée, sont proches du centre du palindrome. La structure de la chromatine de cette région est indiquée dans la partie agrandie de la carte.

Plusieurs sites hypersensibles sont localisés dans cette région. Chaque promoteur en possède un, tout comme chaque origine. Une série d'exactly 5 nucléosomes est située entre les deux origines. Ces origines étant séparées par 1 000 pb et l'espacement des nucléosomes étant exactement de 200 pb, chaque nucléosome occupe donc une position définie. Dans un autre type de *T. pyriformis*, la région centrale entre les deux origines est longue de 1 400 pb et contient 7 nucléosomes.

L'ADNr de *Tetrahymena* constitue un exemple par excellence des nucléosomes en phase. Le fait que les deux versions de l'ADNr diffèrent par des multiples de la

Figure 27.19

La région sans nucléosome de SV40 contient des sites hypersensibles, des régions sensibles et une région d'ADN protégée.



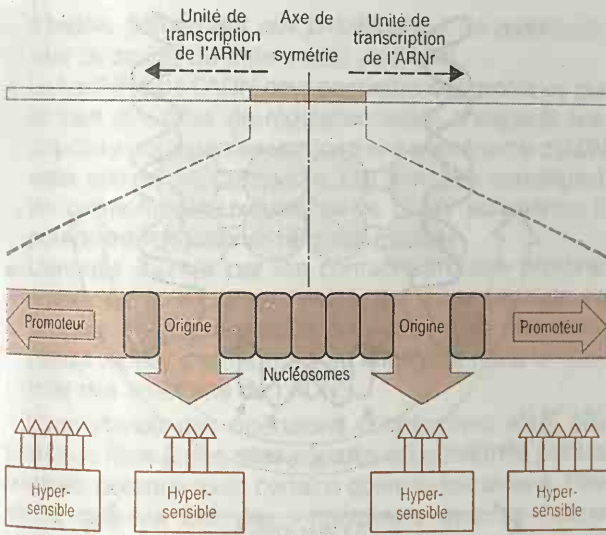


Figure 27.20

La région centrale de l'ADNr de *Tetrahymena* possède une structure chromatinienne étroitement définie avec des sites hypersensibles dans les promoteurs et origines et une série de nucléosomes en phase entre les origines.

répétition de nucléosomes suggère qu'il pourrait exister une pression sélective dans la capacité à compacter la région centrale en un nombre exact de nucléosomes. Les localisations des octamères d'histones sont sans doute définies par les limites des sites hypersensibles de chaque côté, plutôt que par une propriété intrinsèque de la séquence en phase.

Des altérations de structure chromatinienne comparables peuvent également exister dans la région du centromère de levure. Au lieu d'un kinétochore visible dans la levure, il semble probable que le filament de chromatine lui-même puisse être impliqué dans la fonction du centromère. L'étude de la structure de la région centromérique a été réalisée en utilisant un court fragment d'ADN comme sonde dans la technique de marquage indirect d'extrémités. La sonde reconnaît deux sites hypersensibles à l'ADNase I, localisés de chaque côté des régions I et III, 2 des courtes séquences hautement conservées dans l'ADN centromérique dont les séquences sont résumées dans la Figure 25.11. La région d'environ 220-250 pb entre les deux sites hypersensibles est protégée de la digestion par la nucléase ; ainsi le rôle de la région conservée pourrait être de lier les protéines centromériques spécifiques qui forment une structure non nucléosomale.

Une expérience identique de digestion de la chromatine par la nucléase micrococcale permet de mettre en évidence un autre effet. La même région protégée apparaît avec, en outre, une série de nucléosomes en phase qui s'étend de chaque côté. La phase se maintient sur plus de 12 nucléosomes, chacun contenant 160 pb d'ADN.

La mise en phase des nucléosomes se propage-t-elle à partir des séquences centromériques elles-mêmes ou est-ce une propriété des séquences qui sont en phase ? Dans un plasmide qui contient ces séquences mais qui est dépourvu de la région centromérique, la phase est maintenue. Ce résultat suggère que la région centromérique authentique possède deux particularités. Une séquence conservée préside à l'établissement d'une certaine structure non nucléosomale et les séquences bordantes semblent posséder la capacité intrinsèque de maintenir en phase leurs nucléosomes.

Qu'advient-il de l'arrangement des nucléosomes à l'extrémité d'un chromosome ? Lorsqu'ils sont clivés par la nucléase micrococcale, les télomères du cilié *Oxytricha*, présentent une série de bandes répétées à 100, 300, 500, 700 et 900 pb à partir de l'extrémité. Ce type de séparation laisse penser que les 100 dernières pb du télomère sont protégées par un complexe protéique non nucléosomal adjacent à une série de nucléosomes en phase.

Comment les sites hypersensibles (ou autres) se propagent-ils à travers la réplication de l'ADN ? On peut imaginer plusieurs mécanismes. Un changement transmissible pourrait s'effectuer dans l'ADN lui-même, par exemple, une déméthylation. A partir de la double hélice normale la structure de l'ADN pourrait être différente (par l'intermédiaire de protéines permettant à la structure de se reformer après chaque réplication). Les protéines peuvent former un complexe (avec l'ADN double brin) qui ségrège à la réplication et se reforme ensuite.

Trois changements qui apparaissent dans les gènes actifs ont été décrits. Un (ou des) site hypersensible est formé à proximité du promoteur. Les nucléosomes d'un domaine comprenant la région transcrite deviennent plus sensibles à l'ADNase I. L'ADN de cette même région est hypométhylé. Tous ces changements sont peut-être nécessaires à la transcription.

FLEXIBILITÉ STRUCTURALE DANS L'ADN

L'ADN ne devrait pas être conçu en tant que molécule double brin linéaire étirée ; il existe au sein d'une structure topologique définie (nucléosomale ou autre) qui est essentielle à sa fonction. A propos de sa topologie, une

question majeure concerne sa rigidité : *quel est le degré de flexibilité d'un ADN double brin ?* D'une façon plus pratique : *quelle quantité d'énergie requiert la déformation de l'ADN en partant de la topologie d'une molécule double brin linéaire ?*

Nous savons, de par sa compaction dans les virions, les nucléoïdes bactériens et les noyaux eucaryotes (voir Tableau 25.1), que l'ADN peut être fortement condensé dans des conditions physiologiques. Nous aimerions savoir si ces mécanismes de condensation entraînent une déformation continue de l'hélice ou bien des courbures importantes ou des « vrilles ».

Un ADN double brin peut être très fermement maintenu courbé. Une séquence ne contenant que 169 pb peut être fermée sous forme d'un cercle. Un tel cercle possède une propriété intéressante. *Il est attaqué par l'ADNase I à la même périodicité de 10,5 pb, que celle qui caractérise également l'ADN nucléosomal.* La périodicité reflète le fait que « l'extérieur » du double brin est accessible tandis que « l'intérieur » est protégé. Il est possible que la périodicité soit produite quelle que soit la méthode utilisée pour faire adopter à l'ADN double-brin une structure circulaire-fermeture d'une petite molécule de façon covalente, enroulement autour du nucléosome, liaison à d'autres complexes protéiques.

Ceci nous amène à une expérience ingénieuse réalisée avec le répresseur du phage lambda. Souvenons-nous de la Figure 12.19 : les répresseurs lambda se lient de façon coopérative à leurs sites opérateurs. Quand un dimère répresseur se lie à un site opérateur, un second dimère peut se lier de façon encore plus efficace au site opérateur adjacent. Les domaines en N-terminal du répresseur se lient à l'ADN et les domaines en C-terminal sont orientés de telle façon qu'ils permettent un contact entre les dimères adjacents, créant l'effet coopératif.

La Figure 27.21 montre que la liaison coopérative n'est pas limitée aux sites opérateurs adjacents ; elle apparaît quand les deux sites opérateurs sont séparés par un nombre entier (5 ou 6) de tours de la molécule double brin. Elle n'apparaît pas quand les sites sont séparés par 4,5, 5,5 ou 6,5 tours.

Un des deux sites opérateurs possède une forte affinité pour le répresseur ; l'affinité de l'autre est plus faible. La coopérativité dans la liaison réduit la quantité de répresseur nécessaire pour compléter le second site. Quand une liaison coopérative se produit, l'ADN situé entre les deux sites opérateurs devient sensible à l'action de l'ADNase I avec une périodicité de 10,5 pb !

L'interprétation de ces résultats est illustrée par la Figure 27.21. Quand les sites sont séparés par 63 pb (6 tours d'hélice), les dimères répresseurs s'unissent du

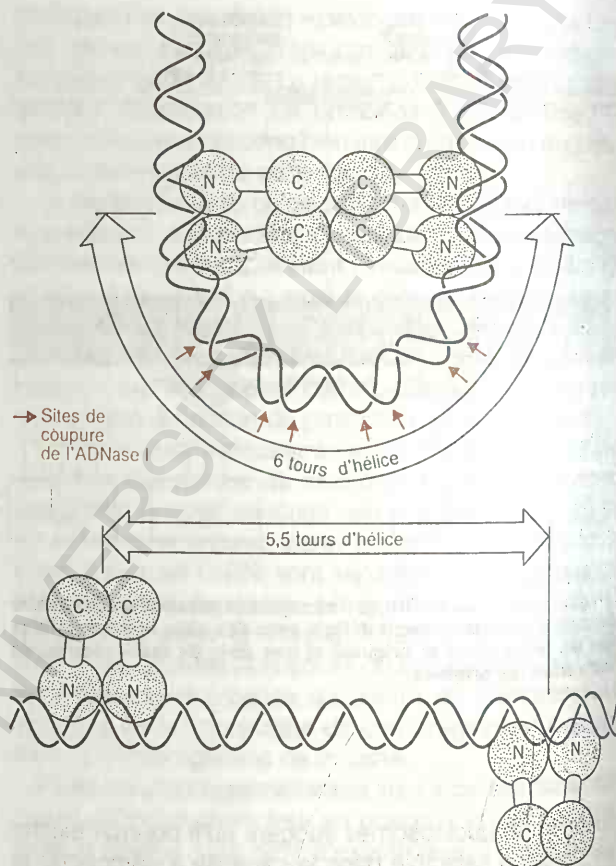


Figure 27.21

Les dimères du répresseur lambda se lient de façon coopérative aux sites opérateurs séparés par un nombre entier de tours d'hélice, mais se lient individuellement à des sites séparés par un nombre entier et demi de tours.

même côté de l'ADN. Les domaines C-terminaux peuvent entrer en contact en courbant l'ADN ; et l'énergie libérée par le contact protéine-protéine est suffisant pour maintenir cette courbure. Mais si ces sites sont séparés par 59 pb (5,5 tours) les dimères répresseurs se lient sur les côtés opposés de l'ADN. Une plus grande quantité d'énergie serait alors nécessaire pour courber (et tordre) l'ADN afin de permettre le contact entre les domaines C-terminaux. Les contacts protéine-protéine ne sont pas suffisants pour passer outre cet obstacle.

Ces résultats suggèrent quelques conclusions importantes concernant les sites de régulation :

- Quand une région régulatrice contient, au sein d'un domaine restreint, de multiples sites séparés, la distance *per se* ou la séquence qui les sépare peut être relativement peu importante si l'on tient compte du fait qu'ils sont séparés par un nombre entier de tours

d'hélice, permettant aux protéines qui les reconnaissent de se lier du même côté de l'ADN.

- La flexibilité de l'ADN peut permettre aux protéines qui se lient aux sites de régulation isolés d'élaborer une structure qui organise une longueur appréciable d'ADN sous une forme compacte. Les modules constituant les promoteurs auxquels se lie l'ARN polymérase II pourraient fonctionner de cette manière.
- L'énergie fournie par les contacts protéine-protéine entre des facteurs se liant à des sites régulateurs séparés peut être utilisée pour construire des complexes ADN-protéine qui sont impliqués dans le contrôle des fonctions de l'ADN.

L'importance des opérations coopératives entre des protéines liées à des sites séparés est confirmée par les résultats obtenus avec certains opérons bactériens. Plusieurs opérons ont des « opérateurs doubles » dans lesquels un répresseur doit se lier à deux sites non adjacents pour prévenir la transcription. Dans l'exemple de l'opéron *deo*, il y a deux opérateurs qui sont séparés par 599 pb (soit 57 tours d'hélice) ; l'un et l'autre doivent être liés pour que la répression soit complète.

Quand la distance entre les sites séparés est grande, il est possible de la faire varier considérablement sans abolir l'effet. Les opérateurs *deo* fonctionnent tout aussi bien qu'ils soient séparés par 798 pb (76 tours) ou par 997 pb (95 tours). Dans l'opéron *ara* les sites de liaison nécessaires à la répression sont éloignés de 200 pb. Des changements de la distance de séparation par des insertions qui correspondent à des nombres entiers de tour d'hélice n'affectent pas la répression. Mais celle-ci est perturbée par des insertions qui ajoutent un demi-tour entre les sites.

Probablement toute séquence d'ADN peut être courbée si la liaison coopérative des protéines libère suffisamment d'énergie. Mais, de plus, certaines séquences d'ADN ont par elle-même la propriété intrinsèque de se courber. Certains fragments d'ADN ont des mobilités électrophorétiques anormales probablement dues à des régions incurvées. Par exemple, l'origine de réplication de l'ADN de lambda semble ainsi adopter une conformation courbe. Dans cet ADN particulier aussi bien que dans un autre, le génome mitochondrial du kinétoplaste de parasites tropicaux, la région incurvée contient une série de répétitions A-T. L'incurvation semble apparaître à la jonction entre chaque répétition (A_{5-6}/T_{5-6}) et l'ADN-B adjacent. Nous ne savons pas quelle est la signification de la courbure de l'ADN de kinétoplaste mais, pour ce qui est de l'origine de lambda, elle intervient probablement dans la reconnaissance par les protéines de réplication et dans la construction de l'appareil de réplication.

HYPOTHÈSES À PROPOS DE LA NATURE DE L'ACTIVATION DES GÈNES

Des sites détectables par l'hypersensibilité aux nucléases apparaissent dans des situations variées : en amont des gènes qui peuvent être transcrits, dans la zone dépourvue de nucléosome et sensible à la nucléase de SV 40 et de polyome, aux origines de réplication dans l'ADN ribosomal de *Tetrahymena*, et au niveau des centromères de levure ; ils peuvent aussi être associés avec la transposition des gènes de type sexuel chez la levure (voir Chapitre 33). Parfois une série de nucléosomes est localisée à des positions fixes à côté du site hypersensible soit parce que le site fournit une frontière, soit parce que ces régions possèdent la propriété intrinsèque de maintenir les nucléosomes en phase.

La diversité des situations associées à la présence de sites hypersensibles suggère que leur existence est le reflet d'un principe général. *Des sites au niveau desquels la double hélice initie une activité particulière sont maintenus libres de nucléosomes.* Dans chaque cas, certaines protéines non histone (inconnues), spécifiques de la fonction particulière du site, modifient les propriétés d'un court segment d'ADN de telle sorte qu'en soient exclus les nucléosomes. Les structures formées dans toutes ces situations ne sont pas nécessairement identiques (excepté naturellement qu'elles créent, par définition, un site hypersensible à l'ADNase I).

Nous pouvons faire l'hypothèse que l'inverse de ce principe est peut-être également vrai : *des gènes dont les régions de contrôle sont organisées en nucléosomes ne peuvent pas être exprimés.* Supposons que la formation de nucléosomes survienne, indépendamment de sa séquence dans toute région d'ADN dont les histones n'ont pas été spécifiquement exclues. Dans ce cas, en absence des protéines régulatrices spécifiques, les promoteurs, les enhancers, ainsi que toute autre région de contrôle, seront organisés par les octamères d'histones en un état sous lequel *ils ne peuvent probablement pas être activés.* (Il n'y a à ce jour aucun argument selon lequel il existe des protéines capables de déplacer les histones de l'ADN.)

Il existe des indices que la désorganisation de la structure chromatinienne est nécessaire à l'activation des gènes. Le facteur de transcription de l'ARNr 5S (TF3A) ne peut pas activer *in vitro* les gènes s'ils sont complexés aux histones. Cependant ce facteur peut former le complexe nécessaire à l'activation avec l'ADN libre après quoi l'addition d'histones n'empêche pas le gène de rester actif. Lorsque le facteur s'est fixé, il demeure sur son site permettant une succession d'initiation de la transcription

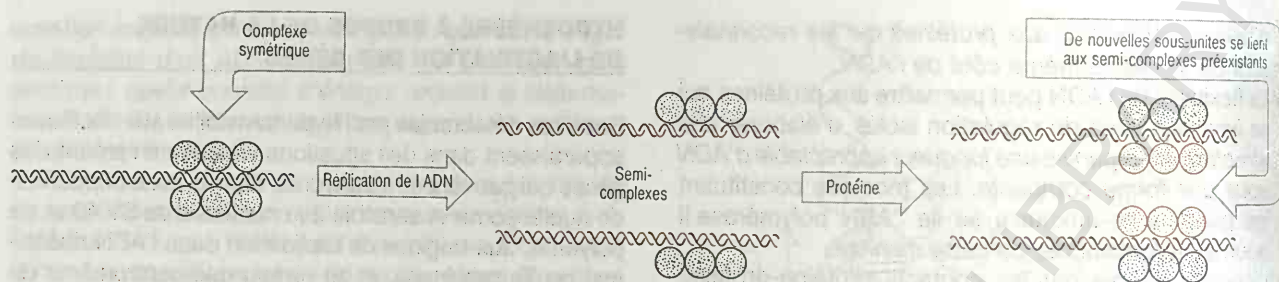


Figure 27.22

Un complexe protéique pourrait se perpétuer sur l'ADN en se répartissant sous forme de demi-complexe lors de la réplication, puis en recrutant de façon coopérative des molécules protéiques supplémentaires pour restaurer le complexe entier.

par les molécules d'ARN polymérase. Le facteur critique dans cette activation peut donc être qui du facteur ou des histones atteindra le premier le site de contrôle.

La réplication fournit une occasion tout à fait évidente de désorganisation de la structure chromatinienne puisque les nucléosomes doivent être transitoirement absents d'au moins un (et peut-être deux) des chromosomes dupliqués. Ceci peut fournir l'occasion pour d'autres protéines de se fixer au site de contrôle empêchant alors la formation secondaire de nucléosomes ? La région exclue correspond alors à un site hypersensible et peut servir de borne à partir de laquelle les nucléosomes s'organisent en phase.

Comment pourrions-nous tester cette idée que la réplication intervient souvent dans la modification des états d'expression de la chromatine ? L'induction de sites hypersensibles dans des gènes qui répondent par exemple aux hormones peut constituer le type de système dont nous avons besoin. Une réponse claire concernant la manière dont se forment les régions actives dépendra cependant du développement d'un système permettant l'activation *in vitro* de la structure chromatinienne.

Une fois qu'un site hypersensible (ou tout autre structure d'activation) s'est formé, il peut se propager à travers le cycle de réplication de façon stable. Ceci pourrait survenir si le facteur avait la capacité de ségréger à chaque réplication. La Figure 27.22 illustre une situation hypothétique dans laquelle une protéine qui se fixe coopérativement à l'ADN forme un complexe qui se scinde lors de la réplication, chacun des demi-complexes étant ensuite restauré sous la forme d'un complexe entier sur les chromosomes fils.

Puisque la reconstitution du complexe est nécessaire à chaque génération, un corollaire de cette hypothèse est que les gènes pourraient être inactivés dans une cellule en division par simple limitation de la quantité du

facteur disponible. Si le complexe en effet se dilue lors des divisions (parce qu'il n'est plus resynthétisé), le complexe ne pourra pas être restauré et les histones formeront des nucléosomes à l'emplacement des sites de contrôle. Ceci pourrait expliquer comment des gènes sont inactivés au cours du développement embryonnaire. La question cependant du mécanisme inactivant un gène dans une cellule qui ne se divise pas reste posée. En effet, des gènes actifs peuvent-ils être inactivés dans des cellules en différenciation terminale ? (c'est-à-dire des cellules qui ont atteint leur stade final de différenciation phénotypique).

Les mécanismes exacts de l'activation de la chromatine (ou de son inactivation) restent trop hypothétiques pour les considérer dès maintenant, mais le sentiment général est que les chromatines actives et inactives ne sont pas en équilibre dynamique. Des événements soudains, de désorganisation sont nécessaires pour convertir l'une en l'autre. Une conséquence de ce type de modèle est que la concentration d'une protéine régulatrice nécessaire à l'initiation d'un événement d'activation est beaucoup plus élevée que celle nécessaire au simple maintien de la chromatine dans sa conformation native.

Des événements intervenant dans l'activation d'un gène ne coïncident pas nécessairement avec l'activation de la transcription. La formation du complexe approprié peut rendre le gène apte à être transcrit mais des phénomènes ultérieurs peuvent être nécessaires pour qu'il soit réellement transcrit. Dans le langage de la biologie du développement, la formation du complexe peut être un événement de *détermination* précédant d'un certain temps l'événement de *différenciation* c'est-à-dire, la véritable activation transcriptionnelle du gène. Un exemple possible de ce phénomène est celui des gènes dont l'expression est contrôlée par les hormones : ils peuvent entrer dans l'état « déterminé » dans les cellules

appropriées, puis être ensuite réellement activés lorsque survient la stimulation hormonale. (Selon ces termes, on pourrait dire que les gènes des procaryotes sont contrôlés par des mécanismes ressemblant à ceux de la différenciation puisque leur activation correspond en fait au réel début de leur transcription plutôt qu'en la création d'un état de disponibilité.)

POUR EN SAVOIR PLUS

L'existence de structures non nucléosomales dans les fibres de chromatine a été mise de jour par **Eissenberg et al.** (*Ann.*

Rev. Genet. **19**, 485-536, 1985). La propagation des sites hypersensibles a été suivie par **Groudine & Weintraub** (*Cell* **30**, 131-139, 1982). Un tour d'horizon des phénomènes impliqués dans l'établissement et dans le maintien des structures de gènes actifs a été fait par **Brown** (*Cell* **37**, 359-365, 1984) ; et les relations structure-fonction de la chromatine ont été explorées par **Weintraub** (*Cell* **42**, 705-711, 1985). La liaison coopérative de protéines à des sites de contrôles séparés a été rapportée par **Hochschild & Ptashne** (*Cell* **44**, 681-687, 1986). Les hypothèses sur la méthylation ont été le sujet d'une revue de **Razin & Riggs** (*Science* **210**, 604-610, 1980). La structure des centromères a été révélée par **Bloom & Carbon** (*Cell* **29**, 305-317, 1982).

PARTIE 9

LE GÉNOME DYNAMIQUE : L'ADN EN MOUVEMENT

PARTIE 9

LE GÉNOME DYNAMIQUE : L'ADN EN MOUVEMENT

Les éléments portés par les chromosomes de maïs... contrôlent l'action du gène et induisent au niveau du site du gène, des modifications héréditaires concernant cette action. Ces éléments furent initialement découverts parce qu'ils ne restaient pas en une position donnée sur le chromosome complémentaire. Ils peuvent apparaître à de nouveaux emplacements et disparaître de ceux établis précédemment. La présence d'un tel élément, sur, ou à côté du locus d'un gène connu, peut affecter le rôle de ce gène. Ceci ne nécessite pas la modification de l'action potentielle des substances géniques sur le locus. C'est pourquoi ces éléments furent appelés éléments de contrôle... On pourrait considérer qu'un élément de contrôle représente une sorte de substance extrachromosomique qui peut se fixer, ou laisser l'empreinte de son influence en quelque sorte, en différentes positions du chromosome complémentaire, et ainsi modifier l'action des substances géniques situées en ces positions. Pourtant, les modes d'action des éléments de contrôle ne suggèrent pas cela. Ils suggèrent plutôt, que les éléments de contrôle sont des composants intégraux des chromosomes eux-mêmes, et qu'ils ont des activités spécifiques et les moyens de les réaliser, comme les gènes eux-mêmes sont supposés avoir... Les transpositions des éléments de contrôle proviennent, soit de quelque mécanisme encore inconnu, ou surviennent lors du processus de duplication du chromosome lui-même, et en sont une conséquence.

Barbara McClintock, 1956

CHAPITRE 28

RECOMBINAISON ET AUTRES MANIPULATIONS TOPOLOGIQUES DE L'ADN

Sans la recombinaison génétique, la composition de chaque chromosome serait irrémédiablement établie, avec ses propres allèles, variant seulement par mutation. La longueur de la cible correspondant au dégât de la mutation serait étendue du gène au chromosome. Les mutations délétères s'accumuleraient, éliminant chaque chromosome (et par ce moyen, annulant toutes mutations bénéfiques ayant pu survenir).

Si l'on mélange les gènes, la recombinaison permet de séparer des mutations favorables et défavorables, et de les étudier comme des entités individuelles, dans un nouveau contexte. Cela procure aux allèles favorables, le moyen de s'échapper et de se développer, et permet d'éliminer un allèle non favorable, sans détruire les autres gènes auxquels cet allèle aurait pu être associé dans le passé. Dans la longue perspective de l'évolution, un chromosome est un oiseau de passage, une association temporaire d'allèles particuliers. La recombinaison est responsable de ce comportement instable.

La recombinaison impliquant une réaction entre des séquences homologues d'ADN est appelée **recombinaison généralisée**. Les enzymes responsables peuvent utiliser *n'importe* quelle paire de séquences homologues

comme substrat (bien que certains types de séquence soient favorisés par rapport à d'autres). La fréquence des recombinaisons n'est pas constante sur tout le génome, mais peut être influencée par la structure du chromosome ; par exemple, le crossing-over peut être réprimé au voisinage de l'hétérochromatine.

Un autre type d'événement aboutit à la recombinaison entre des paires de séquences *spécifiques*. La **recombinaison spécifique de sites** est responsable de l'intégration du génome d'un phage dans le chromosome bactérien. Elle implique les séquences spécifiques des ADN de phage et de bactérie. Une courte région d'homologie à l'intérieur de ces séquences est nécessaire, mais non suffisante, pour que se produise la recombinaison. Les enzymes responsables de cet événement ne peuvent pas recombiner d'autres paires (homologues ou non homologues) de séquences, mais agissent seulement sur la paire de séquences particulière du phage et de la bactérie.

Une autre sorte d'événement permet à une séquence d'ADN de s'insérer dans une autre sans relation avec l'homologie de la séquence. La **recombinaison répllicative** permet à certains éléments de se déplacer d'un point d'un chromosome à un autre. L'élément mobile d'ADN

est copié, puis la réplique est transférée ailleurs. Ce sera le sujet du Chapitre 29.

LA RECOMBINAISON NÉCESSITE UN ACCOLEMENT D'ADN DOUBLE BRIN HOMOLOGUES

La recombinaison entre chromosomes implique un échange physique de régions (voir Figure 1.11). La structure créée par cet échange est visible lors de la méiose sous la forme d'un chiasma (voir Figure 1.10). Le chiasma représente le résultat d'une **cassure et d'une soudure**, lors desquelles deux chromatides non sœurs (contenant chacune un double brin d'ADN) ont été cassées puis reliées l'une à l'autre. La recombinaison se fait avec précision, entre des séquences correspondantes, de telle sorte qu'aucune paire de bases ne soit ajoutée ou retirée aux chromosomes recombinants.

L'échange de séquences représente l'acte de recombinaison lui-même. Sa description nécessite le déroulement de la première étape cruciale : *deux molécules d'ADN*

double brin homologues doivent être amenées en étroit contact, de telle sorte que les séquences correspondantes puissent s'échanger. La nature des événements impliqués dans cette étape peut être quelque peu différente chez les eucaryotes et les procaryotes.

L'ADN des eucaryotes est étroitement empaqueté dans les structures distinctes des chromosomes. Comment les molécules d'ADN se juxtaposent-elles ? Le contact entre une paire de chromosomes parentaux intervient tôt lors de la méiose. Le processus est appelé **synapse ou appariement chromosomique**. Les chromosomes homologues (chacun se composant réellement de deux chromatides sœurs issues de la réplication précédente) s'approchent l'un de l'autre. Ils sont associés latéralement et forment un **complexe synaptique** ayant une structure caractéristique pour chaque espèce, malgré de larges variations dans les détails entre les espèces.

Un exemple de complexe synaptique est représenté sur la **Figure 28.1**. Chaque chromosome à cette étape apparaît telle une masse de chromatine bordée d'un **élément latéral** (qui, dans ce cas, a une structure striée). Les deux éléments latéraux sont séparés l'un de l'autre

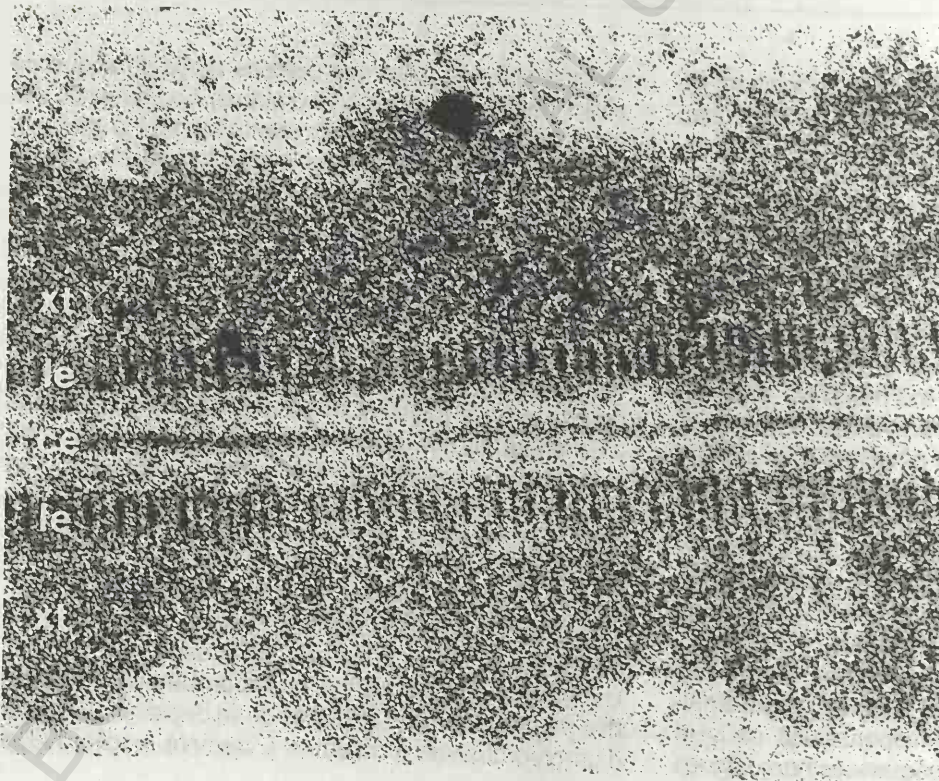


Figure 28.1

Le complexe synaptique de *Nectella* montre que l'appariement des chromosomes ne juxtapose pas les ADN homologues.

xt indique la chromatine ; le indique les éléments latéraux, mesurant chacun à peu près 50 nm de diamètre ; ce indique l'élément central d'environ 18 nm de diamètre. La distance entre les éléments latéraux est d'environ 120 nm.

Photographie aimablement fournie par M. Westergaard et D. Von Wettstein.

par un **élément central**. Le triplet des brins parallèles et denses est situé dans un plan unique qui s'incurve et tourne le long de son axe. La distance entre les chromosomes homologues est considérable en termes moléculaires, plus de 200 nm (le diamètre de l'ADN est de 2 nm).

L'apparition du complexe synaptique coïncide avec le moment présumé du crossing-over, bien qu'il ne soit pas certain que la recombinaison survienne au stade de la synapse. Un problème majeur dans la compréhension du rôle du complexe est que, bien qu'alignant des chromosomes homologues, il est loin de mettre en contact des molécules d'ADN homologues.

Le seul chaînon visible entre les deux côtés du complexe synaptique est produit par des structures sphériques ou cylindriques observées chez les champignons et les insectes. On les trouve dans tout le complexe et ils sont appelés **nœuds** ou **nodules de recombinaison**. Ils surviennent avec la même fréquence et selon la même distribution que les chiasmas. Leur nom fait penser qu'il pourrait s'agir des sites de recombinaison.

Au stade suivant de la méiose, les chromosomes se débarrassent du complexe synaptique ; ensuite, les chiasmas deviennent visibles, tels des points auxquels les chromosomes s'accrochent. Cela annonce le moment de l'échange génétique. Plus tard, au cours de la méiose, les chiasmas peuvent se déplacer vers les extrémités des chromosomes. Cette souplesse suggère qu'ils peuvent représenter un vestige de la recombinaison plutôt que le véritable intermédiaire.

Dans le système procaryote, l'ADN lui-même (quoique dans le nucléoïde), pose de nombreux problèmes. Il n'y a pas d'indication visible des événements responsables de la reconnaissance de l'homologie, mais ils impliquent probablement uniquement les régions qui participent à la recombinaison par chevauchement (cross-over). Cette réaction de reconnaissance peut représenter une part plus ou moins importante du mécanisme de recombinaison lui-même.

CASSURE ET SOUDURE IMPLIQUENT UN DOUBLE BRIN HÉTÉROLOGUE D'ADN

Contrastant avec la base inconnue de l'aptitude des séquences double brin d'ADN homologues à se reconnaître l'une l'autre, un mécanisme d'une grande précision — appariement — existe, celui de la reconnaissance entre simples brins complémentaires. Ceci est utilisé dans la formation des intermédiaires de recombinaison. Un, parmi plusieurs modèles de réaction, est illustré par la Figure 28.2.

Le processus s'amorce par des cassures aux points correspondants des brins homologues de deux ADN double brin appariés. La cassure permet le mouvement des extrémités libres créées par les brèches. Les deux chaînes quittent leur partenaire, se croisent pour s'apparier avec le brin complémentaire de l'autre double brin d'ADN.

L'échange réciproque crée une liaison entre les deux ADN double brin. Initialement, ceci ne repose que sur des liaisons hydrogène ; à un certain moment, elle est rendue covalente par la soudure des brèches sur les sites d'échange (la soudure peut intervenir à un stade plus tardif que celui montré sur la figure). La paire de double-brins connectée est appelée **molécule jointe**.

Au niveau du site de recombinaison, chaque double brin possède une région constituée par un brin de chacune des molécules d'ADN parental. Cette région est appelée **ADN hybride** ou **ADN double brin hétérologue**. La représentation de l'édifice montre (d'une manière surprenante) peu d'inhibition stérique de la formation des régions double brin hétérologues réciproques entre les molécules d'ADN appariées ; ainsi, de ce fait, toutes les bases peuvent demeurer des bases appariées (c'est-à-dire soit dans le double brin parental, soit dans les régions double brin hétérologues). L'existence de régions simple brin étendues n'est pas nécessaire.

Quelle est la longueur minimum de la région nécessaire pour établir une connexion entre les double brins recombinants ? Des expériences lors desquelles de courtes séquences homologues de plasmides ou de phages ont été introduites dans des bactéries suggèrent que la vitesse de recombinaison est notablement réduite si la région homologue est inférieure à 75 pb. Cette distance est sensiblement plus grande que les quelque 10 pb nécessaires à l'association entre les régions simple brin complémentaires, ce qui suggère que la réaction requiert des conditions supplémentaires pour la fusion parfaite des compléments.

Lorsque l'échange de brin est initié, il peut se propager le long du double brin. Cette capacité est appelée **déplacement de la fourche**. Les fourches, dans une structure d'ADN, peuvent impliquer de simples ou doubles brins ; chaque type peut se déplacer. La Figure 28.3 donne un exemple de déplacement de fourche simple brin. Le point d'embranchement peut migrer dans l'une ou l'autre direction à mesure qu'un brin est remplacé par l'autre.

Le déplacement de la fourche est important pour des raisons à la fois théoriques et pratiques. Sur le plan des principes, il confère une propriété dynamique aux struc-

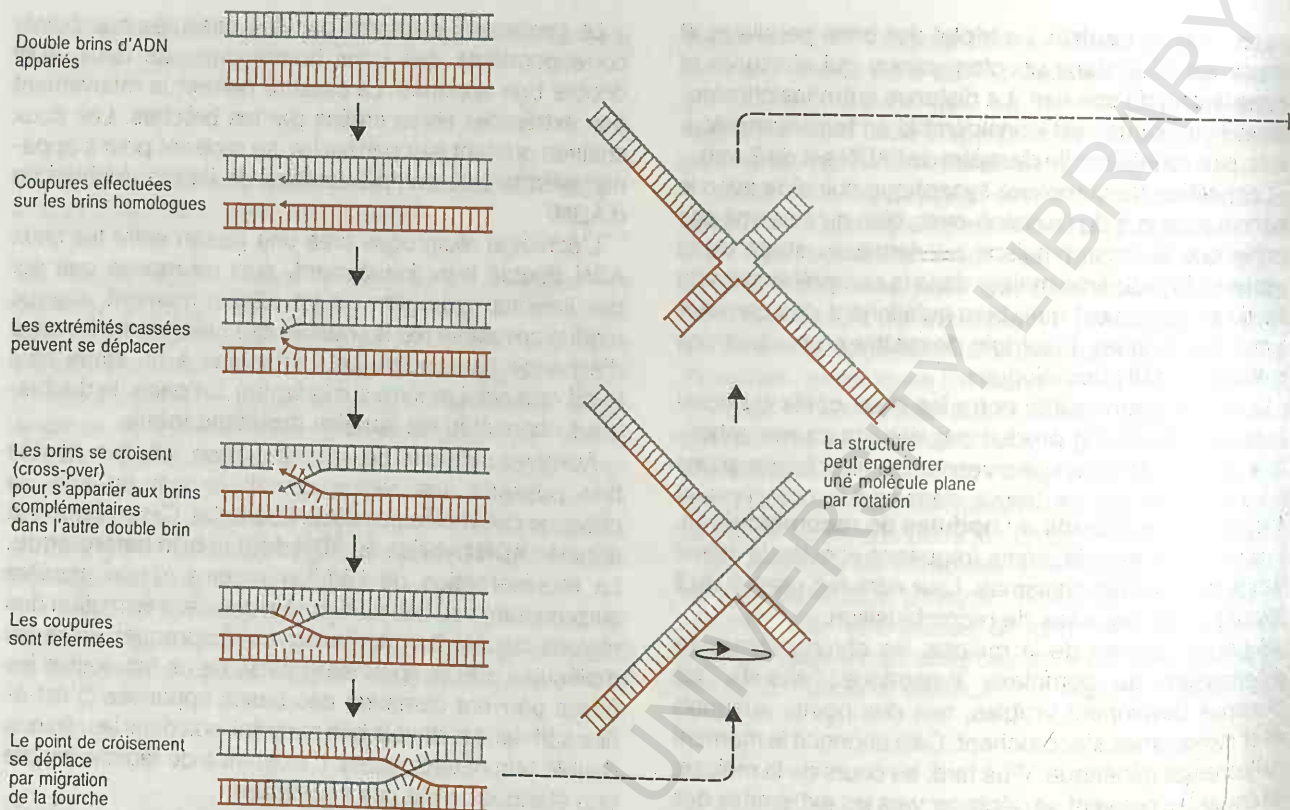


Figure 28.2

La recombinaison entre deux ADN double brin appariés peut être amorcée par un échange réciproque simple brin, poursuivi par le déplacement de la fourche, et résolue par coupure.

tures recombinantes. En pratique, son existence signifie que le point de raccordement ne peut être déterminé en examinant une molécule *in vitro* (parce que le raccordement peut avoir migré depuis l'isolement de la molécule).

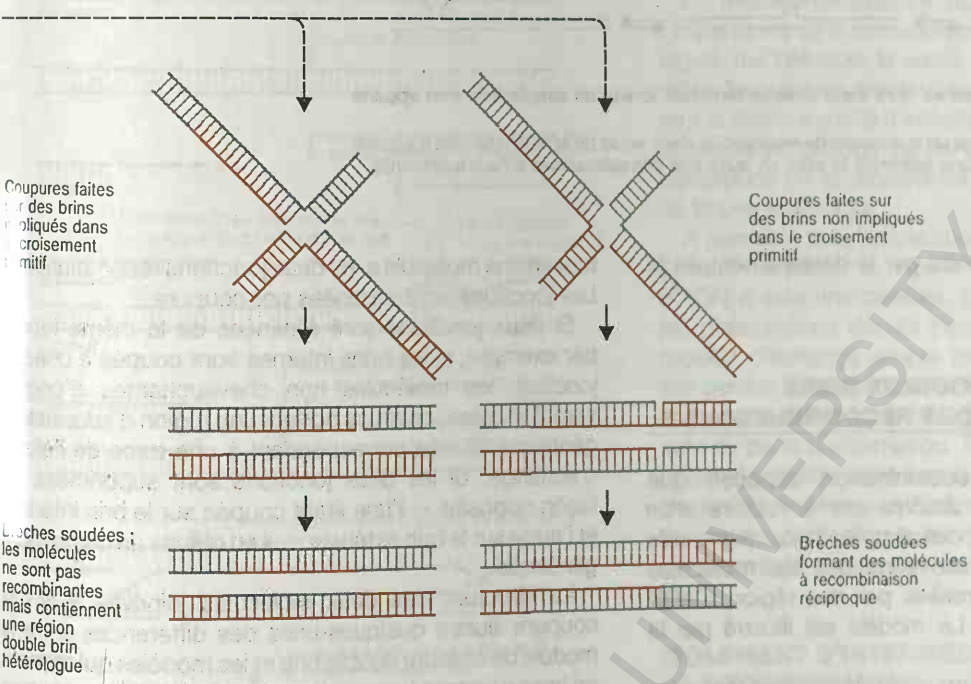
Le même type de mouvement pourrait permettre, au point de croisement d'un intermédiaire de recombinaison, de se déplacer dans l'une ou l'autre direction. La vitesse de déplacement de la fourche est incertaine, mais estimée à environ 30 pb/sec. Bien que le point de raccordement exact puisse bouger, la vitesse est certainement inadaptee pour permettre la formation de régions étendues de double brins hétérologues d'ADN dans des conditions normales. Tout déplacement important de fourche *in vivo* doit donc être catalysé par une enzyme de recombinaison.

Lorsqu'une recombinaison implique des molécules d'ADN double brin, une manipulation topologique est nécessaire ; ou bien l'ADN double brin doit pouvoir

tourner librement, ou bien un phénomène équivalent doit permettre de relâcher la contrainte topologique (voir plus loin).

La molécule jointive formée par l'échange de brins doit être **décomposée** en deux molécules double brin séparées. Ceci nécessite une nouvelle paire de coupures. Les possibilités de cette réaction sont représentées à droite de la Figure 28.2, en considérant la molécule plane obtenue par la rotation d'un des double brins de l'intermédiaire de recombinaison. Les conséquences varient selon la paire de brins qui est coupée.

Si les coupures sont faites dans la paire de brins qui initialement n'a pas été coupée (la paire qui n'est pas à l'origine de l'échange de brins), les quatre brins originaux auront été clivés. Cela libère les molécules d'ADN recombinantes. Le double brin d'un des ADN parentaux est lié de façon covalente au double brin de l'autre ADN parental, via un segment d'ADN double brin hétérologue. Il y a eu



une recombinaison conventionnelle entre les marqueurs situés d'un côté ou de l'autre de la région double brin hétérologue.

Si les deux *mêmes* chaînes qui ont été impliquées lors de l'ouverture d'origine sont à nouveau clivées, les deux brins restants demeurent intacts. La coupure libère les doubles brins parentaux, si ce n'est que chacun porte une trace de l'événement sous la forme d'un fragment d'ADN double brin hétérologue.

Ces résolutions alternatives de la molécule jointe établissent le principe *qu'un échange de brins entre deux ADN double brin laisse toujours après lui une région d'ADN double brin hétérologue, mais que l'échange peut être ou ne pas être accompagné d'une recombinaison des régions bordantes.*

Plusieurs variations sur l'utilité de l'ADN double brin hétérologue peuvent être envisagées. L'échange de brin pourrait débiter par seulement un brin et non par deux ;

le déplacement de la fourche pourrait être responsable de la réaction qui déplace l'autre brin. La réaction pourrait survenir à la suite du déroulement et de l'appariement en croix des brins, sans aucune ouverture (bien que ceci impose des besoins topologiques plus importants).

Dans l'exemple de la Figure 28.2, l'ADN double brin hétérologue a la même longueur dans chaque molécule. Des études génétiques sur le champignon (voir plus loin) suggèrent que dans les recombinaisons réciproques de chromosomes, les régions d'ADN double brin hétérologues peuvent être différentes. Ceci peut être expliqué par une modification du modèle dans lequel un nouveau segment d'ADN est synthétisé pour remplacer une partie de l'un des brins s'échangeant.

Tous ces facteurs affectent les détails du mécanisme et l'ordre des événements, mais les modèles reposent sur le même principe général : la formation d'un intermédiaire impliquant un ADN double brin hété-



Figure 28.3

La migration de la fourche peut se faire dans chaque direction lorsqu'un simple-brin non apparié se substitue à un brin apparié.

Cette structure peut être créée par un événement de renaturation dans lequel un ADN simple-brin s'apparie avec un brin complémentaire à une extrémité et avec un autre brin complémentaire à l'autre extrémité.

rologue qui peut être étendu par le déplacement de la fourche.

LES CASSURES DES DOUBLES BRINS PROVOQUENT-ELLES DES RECOMBINAISONS ?

Un type courant de recombinaison suppose que l'échange génétique est amorcé par la rupture d'un double brin. Ensuite, l'un des simples brins migre vers l'autre double brin, de telle sorte que les molécules double brin se trouvent reliées par une région d'ADN double brin hétérologue. Le modèle est illustré par la Figure 28.4.

La recombinaison est amorcée lorsqu'une endonucléase provoque une rupture double brin dans une chromatide, le *receveur*. L'entaille s'élargit en une brèche, probablement sous l'action d'une exonucléase. L'exonucléase ronge un brin d'un côté ou de l'autre de la brèche, engendrant des extrémités 3' simple brin. L'une des extrémités libres 3' envahit ensuite une région homologue de l'autre double brin *donneur*. La formation d'ADN double brin hétérologue produit une boucle D sur laquelle un brin du donneur double brin est déplacé. La boucle D est allongée par une synthèse de réparation, l'extrémité 3' libre servant d'amorce.

Finalement, la boucle D devient suffisamment grande pour correspondre à la longueur entière de la brèche sur la chromatide *receveuse*. Lorsque le simple brin expulsé atteint le côté le plus éloigné de la brèche, les séquences simple brin se ressoudent. Ainsi, il y a un ADN double brin hétérologue sur l'un ou l'autre côté de la brèche, et la brèche elle-même est représentée par la boucle D simple brin.

L'intégrité du double brin de la région éviscée peut être restaurée par une synthèse de réparation utilisant comme amorce l'extrémité 3' du côté gauche de la brèche. Finalement, la brèche a été réparée par deux procédés distincts de synthèse d'ADN simple brin.

Le déplacement de la fourche transforme cette struc-

ture en une molécule avec deux jonctions recombinantes. Les jonctions sont éliminées par coupure.

Si deux jonctions sont éliminées de la même façon, par exemple, si les brins internes sont coupés à chaque jonction, les molécules non chevauchantes d'origine seront libérées, chacune portant une région d'information génétique altérée correspondant à une trace de l'étape d'échange. Si les deux jonctions sont supprimées de façon opposée — l'une étant coupée sur le brin intérieur et l'autre sur le brin extérieur — il en résulte un croisement génétique.

La structure des deux molécules jointives avant la coupure illustre quelques-unes des différences entre le modèle de cassure double brin et les modèles qui mettent en jeu seulement les simples brins. De l'ADN double brin hétérologue a été formé à chaque extrémité de la région impliquée dans l'échange. Entre les deux fragments d'ADN double brin hétérologue se trouve la région correspondante à la brèche, qui possède maintenant la séquence de l'ADN donneur dans les deux molécules.

Ainsi, l'arrangement des séquences d'ADN double brin hétérologue est asymétrique et une partie d'une molécule a été convertie en la séquence de l'autre (c'est pourquoi la chromatide initiatrice est appelée le *receveur*). Les modèles d'échange simple brin fournissent le pronostic inverse, à savoir que la chromatide initiatrice serait le donneur de l'information génétique. Les résultats chez la levure confirment cette prédiction du modèle des coupures double brin que l'initiation est associée à la réception d'information génétique.

Ce modèle permet de faire des pronostics au sujet des parentés entre les sites impliqués lors de l'échange génétique qui correspondent bien aux résultats obtenus en analysant les systèmes propres aux champignons (voir plus loin). Un des arguments principaux est que les cassures double brin sont connues pour être impliquées dans certains phénomènes de pseudo-recombinaison chez la levure (voir Chapitre 33). Lorsqu'un plasmide porteur d'une cassure double brin est introduit dans une levure, il stimule la recombinaison, et, bien plus, peut

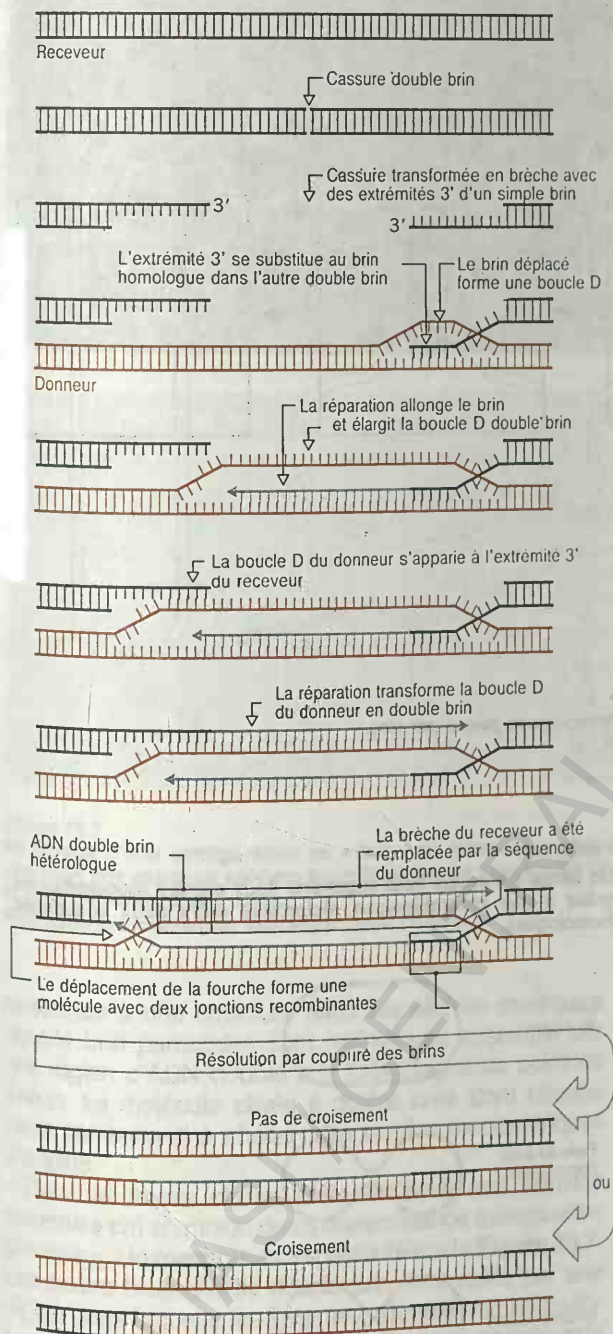


Figure 28.4
La recombinaison peut être initiée par une cassure double brin suivie de la formation d'extrémités 3' simple brin, dont l'une migre vers un double brin homologue.

donner naissance à des éléments dans lesquels une brèche a été réparée.

Le modèle de cassure double brin ne diminue pas l'importance de la formation de l'ADN double brin hétérologue, qui demeure la seule voie plausible d'interaction entre deux molécules double brin. Cependant, en reportant la responsabilité d'initiation de la recombinaison des cassures simple brin sur les cassures double brin, notre conception sur la capacité de la cellule à manipuler l'ADN se trouve influencée.

A première vue, l'implication des cassures double brin semble surprenante. Après tout, lorsqu'une molécule d'ADN a subi une cassure, c'est irréversible. Comparer les phénomènes décrits Figures 28.2 et 28.4 ; dans le modèle d'échange simple brin, aucune information n'a été perdue à aucun moment ; mais dans le modèle de cassure double brin, le clivage initial est immédiatement suivi de perte d'information. Toute erreur dans la récupération de l'information serait fatale. D'un autre côté, la possibilité de récupérer l'information perdue en la resynthétisant à partir d'un autre double brin, procure un filet de sauvetage majeur pour la cellule.

ISOLEMENT D'INTERMÉDIAIRES DE RECOMBINAISON

Une propriété de l'ADN circulaire permet une approche pour isoler des intermédiaires de recombinaison. Nous avons considéré depuis longtemps les doubles brins d'ADN recombinant comme des molécules linéaires, mais beaucoup de génomes (en particulier dans le cas de virus et de plasmides) sont circulaires.

La Figure 28.5 illustre les conséquences d'une recombinaison réciproque entre deux sites homologues de deux molécules d'ADN circulaire. La partie agrandie de la figure montre que la structure au point de jonction des recombinants est identique à celle représentée sur la Figure 28.2 ; la seule différence est que les extrémités de chaque double brin d'ADN parental se rejoignent.

Cette structure possède un aspect topologique important. Sa résolution par la voie de la recombinaison donne naissance à un cercle dimérique, constitué des deux séquences parentales d'origine jointes en position tête-à-queue.

Que cela signifie-t-il pour le style de vie des éléments d'ADN circulaire ? S'il participe à des événements de recombinaison, le génome circulaire d'origine monomérique ne peut se maintenir que grâce à des couples de recombinaisons. Une recombinaison réciproque unique donne toujours naissance au dimère tel qu'on le voit sur la Figure 28.5. Ce dimère peut revenir à la forme morio-

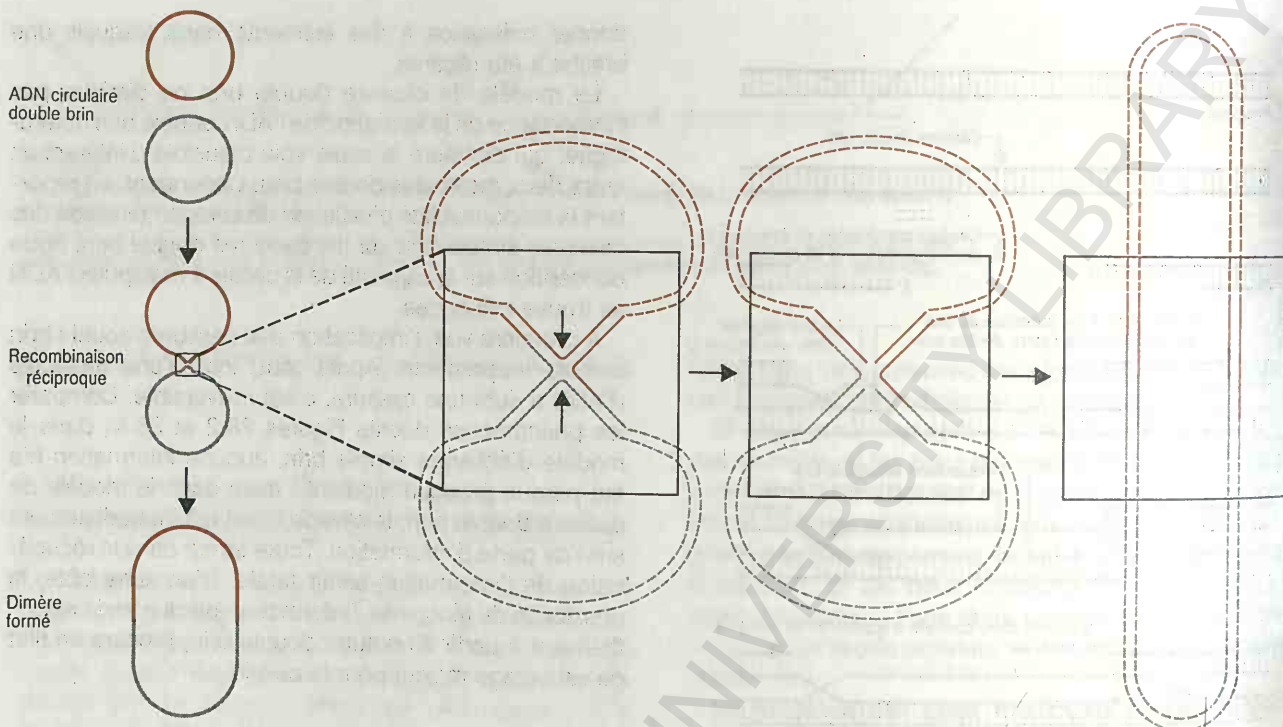


Figure 28.5

La recombinaison réciproque entre deux ADN double brin circulaires homologues passe par une configuration intermédiaire en « huit » pour donner un cercle dimérique.

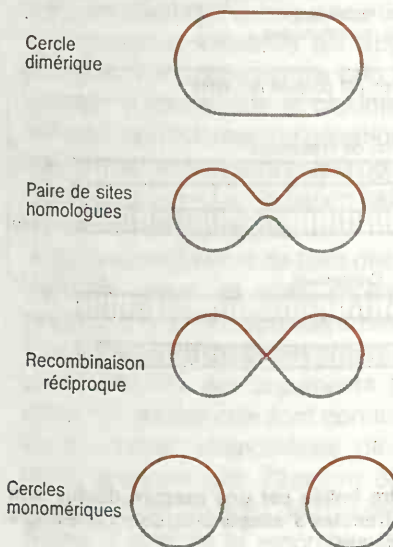
mérique grâce à une seconde recombinaison impliquant n'importe quelle paire de séquences homologues, comme le montre la Figure 28.6. Dans une population quelconque de génomes circulaires, nous pouvons donc nous attendre à trouver des cercles multimériques s'il survient une recombinaison.

Une conséquence pratique de la recombinaison entre ADN circulaires, est que l'intermédiaire de recombinaison devrait prendre la configuration d'un huit, comme le montre la partie agrandie de la Figure 28.5. Peut-on isoler de telles molécules ? Bien que l'on puisse les identifier par la microscopie électronique, leur contour net ne permet pas de les reconnaître sans équivoque comme étant des intermédiaires de recombinaison ; par exemple, un huit peut représenter deux cercles monomériques imbriqués.

La nature de la forme huit peut être reconnue en clivant une molécule isolée par une enzyme de restriction qui coupe chaque cercle monomérique une fois seulement. Les deux parties du huit alors se sépareront *excepté s'ils sont unis par une liaison covalente*. Si la forme huit est un intermédiaire de recombinaison, le clivage doit donner

Figure 28.6

Un cercle dimérique peut engendrer deux cercles monomériques grâce à une recombinaison réciproque entre deux séquences homologues.



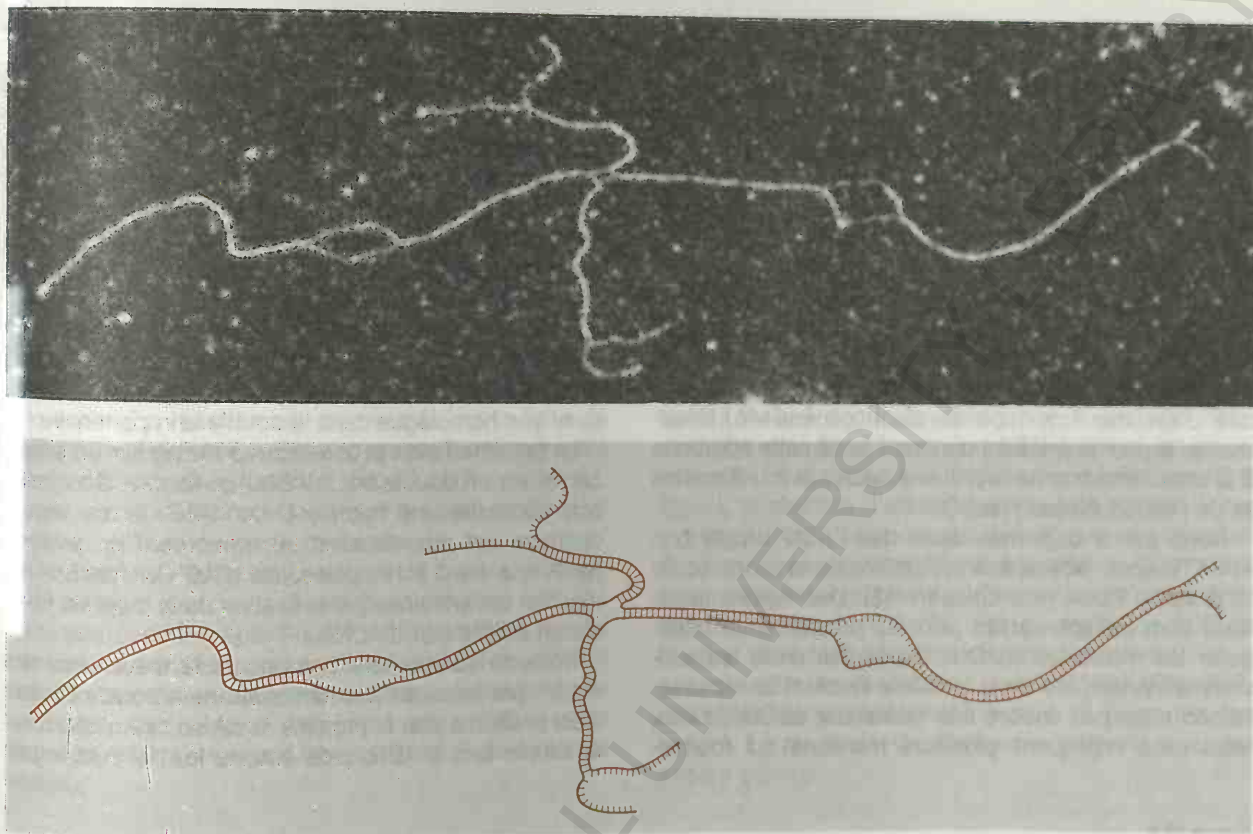


Figure 28.7
Le clivage d'une configuration en « huit » au niveau des sites homologues sur chaque cercle engendre une structure *chi* dans laquelle les quatre bras double brin sont maintenus unis au niveau d'un site d'échange des brins.

Photographie aimablement fournie par David Dressler.

naissance à une structure dans laquelle les molécules double brin parentales sont maintenues ensemble par une région d'ADN double brin hétérologue au point de fusion. La molécule clivée a quatre bras dont chaque paire correspond à chacun des cercles monomériques d'origine.

Dans la forme clivée, cette molécule est appelée **structure chi** en raison de sa ressemblance avec la lettre grecque χ . Un exemple est représenté sur la **Figure 28.7**. Les quatre bras d'ADN double brin sont reliés par une région dans laquelle les fragments double brin hétérologue ont été séparés de leurs simples brins constitutants (certainement par les conditions de préparation). Cette structure correspond exactement à l'intermédiaire prévu dans les Figures 28.2 et 28.5 ; son isolement démontre directement l'existence d'ADN double brin hétérologue *in vivo*.

De telles observations offrent une approche de purification des enzymes de recombinaison. En principe, des extraits acellulaires peuvent être testés pour leur capacité à engendrer des cercles dimériques à partir de cercles monomériques. Malheureusement, il n'a pourtant pas encore été possible de diviser ce système en ses composants et d'utiliser *in vitro* des tests de complémentation, selon des voies d'approche identiques à celles développées précédemment pour la réplication de l'ADN (voir Chapitre 14).

L'ÉCHANGE DE BRIN EST FACILITÉ PAR LA PROTÉINE RecA

Nous connaissons très peu de choses au sujet du mécanisme responsable de la recombinaison en général.

Les enzymes impliquées dans certaines recombinaisons spécialisées ont été caractérisées, et certaines activités topologiques s'y rapportant décrites, mais il a été difficile d'identifier les enzymes qui assument des recombinaisons homologues. Jusqu'à ce que nous disposions d'un dosage fiable qui puisse être utilisé pour purifier les composants nécessaires à la recombinaison ; il est peu probable que nous déterminions toutes les enzymes impliquées.

Une approche plus restreinte consiste à étudier les propriétés des enzymes bactériennes dans la recombinaison, lorsque surviennent des mutations *rec*⁻ dans leurs gènes. Le phénotype mutant *Rec*⁻ correspond à l'incapacité d'assumer une recombinaison généralisée. Néanmoins, le peu de produits connus a limité cette approche à la caractérisation de la protéine RecA aux multifacettes et de l'endonucléase RecBC.

Nous avons déjà mentionné que l'ADN simple brin et l'ATP sont nécessaires à l'activité de la protéase de la protéine RecA (voir Chapitre 15). Les mêmes substrats sont indispensables pour lui permettre de manipuler les molécules d'ADN. La relation entre les activités enzymatiques de la protéine RecA et la recombinaison n'est pas encore très clairement établie *in vivo*, mais elles impliquent plusieurs réactions qui fournis-

sent des paradigmes utiles de mécanismes de recombinaison.

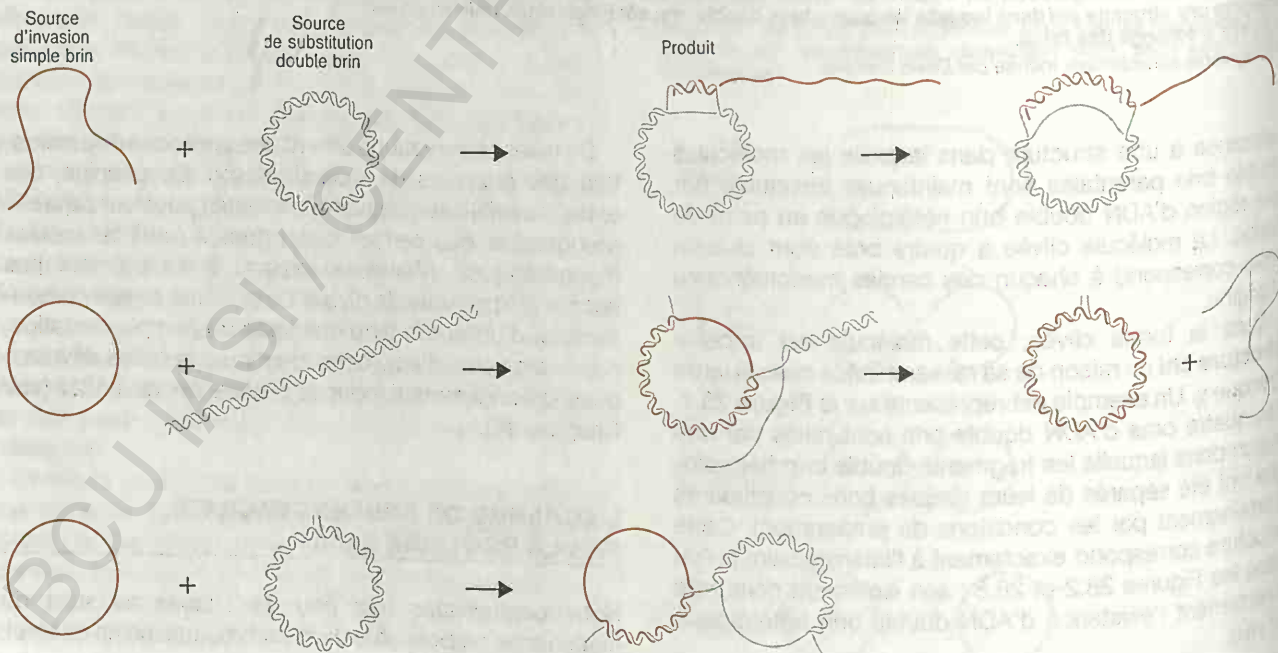
La protéine RecA favorise l'appariement entre un ADN simple brin et son complément dans un double brin superenroulé circulaire. Le simple brin envahit le cercle pour déplacer le partenaire d'origine de son complément, formant ainsi une boucle D, comme le montre la première ligne de la Figure 28.8. Cette réaction est appelée **substitution simple brin** ou **assimilation simple brin**. L'assimilation simple brin est potentiellement apparentée au modèle de formation de double brin hétérologue décrit en Figure 28.2. Le modèle de recombinaison évoque deux invasions, chaque simple brin déplaçant réciproquement son homologue dans le double brin partenaire.

La protéine RecA peut se fixer à la fois sur un simple brin et sur un double brin d'ADN. La réaction d'assimilation nécessite une hydrolyse de l'ATP, et sa vitesse optimum est atteinte avec un monomère de protéine RecA tous les 5 à 10 nucléotides d'ADN simple brin. La réaction est entraînée par la fixation de la protéine RecA sur un simple brin d'ADN.

Nous ne comprenons pas encore le mécanisme des étapes préliminaires à la formation de la boucle D, mais il est probable que la protéine RecA se fixe d'abord sur le simple brin d'ADN, puis ensuite sur le double brin

Figure 28.8

La protéine RecA favorise l'assimilation des simples brins envahissant l'ADN double brin, aussi longtemps que l'un des brins de la réaction possède une extrémité libre.



d'ADN, pour rechercher un complément à la région simple brin. La question principale au sujet de cette opération est de savoir comment l'ADN double brin est dissocié pour permettre l'examen de son éventuelle complémentarité.

Lors de la réaction de la boucle D, le déplacement du brin homologue circulaire n'est que partiel ; la structure du cercle doublé brin covalent voisin peut imposer des contraintes topologiques à la progression de la réaction.

Une réaction plus importante peut se produire si nous inversons les structures des molécules double et simple brin. Une molécule circulaire simple brin peut réagir avec une molécule double brin homologue, pour déplacer un brin entier. La réaction est représentée sur la ligne du milieu de la Figure 28.8. Elle commence à une extrémité du double brin et se poursuit jusqu'à l'autre extrémité, formant un double brin hétérologue qui peut représenter plusieurs milliers de paires de bases.

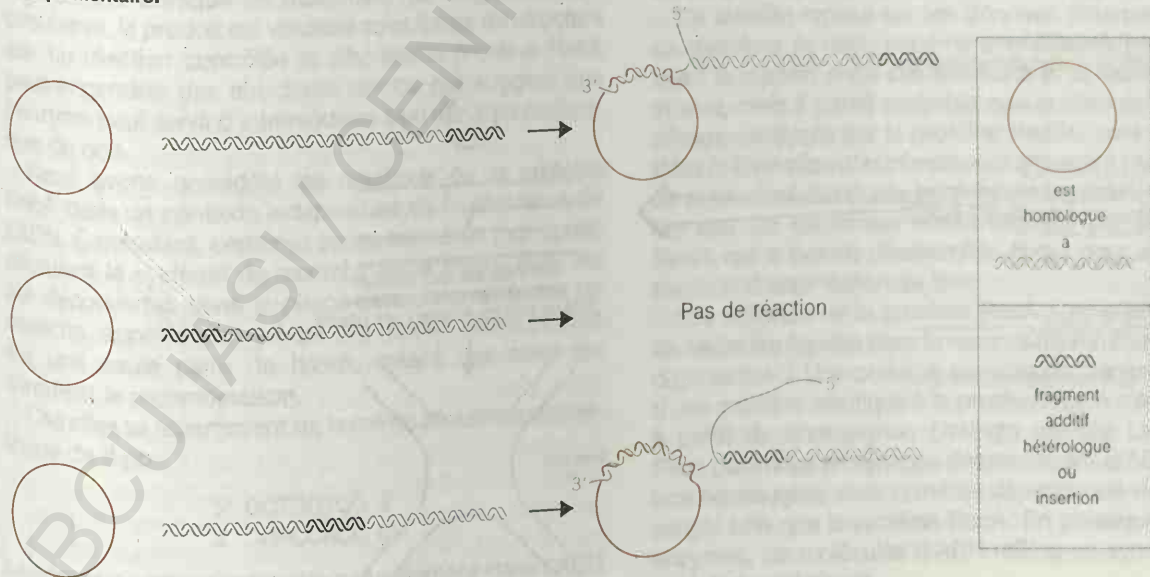
Ce couple de substrats a permis de poser des questions au sujet des mécanismes de la réaction. Va-t-elle dans une direction spécifique ? La Figure 28.9 montre que l'on peut vérifier ceci en utilisant un ADN double brin linéaire complémentaire du simple brin, seulement à une extrémité. Ce substrat est formé par addition d'un fragment d'ADN étranger à une extrémité d'une molécule linéaire.

Quand le fragment est situé à l'extrémité 5' du brin complémentaire — celui qui doit s'apparier avec le simple brin étranger — la réaction se déroule normalement (ligne supérieure de la figure). Mais lorsque l'extrémité 3' du brin complémentaire est constituée par l'ADN étranger, la réaction n'a pas lieu (ligne centrale sur la figure). Les molécules interagissent, mais ensuite se dissocient, car l'assimilation simple brin peut se faire seulement selon une polarité constante, à partir de l'extrémité 3' du complément vers l'extrémité 5', ou, dans les conditions du brin étranger, de 5' en 3'.

Que se passe-t-il lorsque deux séquences sont partiellement complémentaires mais ne s'apparient pas parfaitement ? Quel degré de divergence la protéine RecA peut-elle accepter ? Lorsqu'un ADN étranger est inséré au centre d'un double brin linéaire (ligne inférieure de la figure), la réaction d'assimilation du brin s'arrête à son extrémité. Ainsi, sa poursuite, autant que son initiation, dépend d'un appariement complémentaire. Cependant, l'assimilation de brin peut avoir lieu en dépit d'un modeste défaut de complémentarité. Par exemple, les ADN de phages fd et M13 sont homologues à environ 97 % et peuvent subir une assimilation de brin. Mais l'ADN ϕ X174 et l'ADN G4, qui sont homologues à 70 %, ne peuvent le faire. Le maximum de divergence tolérable n'est pas encore connu.

Figure 28.9

La protéine RecA peut assimiler un simple brin circulaire dans un double brin, seulement s'il existe une extrémité complémentaire 3' avec laquelle le simple brin peut s'apparier. Si un ADN étranger est présent dans le double brin, la réaction d'assimilation s'arrête à l'extrémité de la séquence complémentaire.



LA PROTÉINE RecA ET LES CONDITIONS DE RECOMBINAISON

Le mécanisme d'action de la protéine RecA dans la stimulation du déplacement de la fourche est suggéré par sa capacité à s'agréger en longs filaments avec l'ADN simple brin. En présence d'un cofacteur nucléotidique, la protéine RecA peut s'autopolymériser, ou peut s'incorporer à l'ADN double brin sous forme de filaments. Ces filaments sont plus longs que le double brin d'origine, probablement à cause d'un déroulement de l'ADN.

La présence de SSB (single-strand binding protein = protéine se fixant sur un simple brin) stimule la réaction, en apportant la certitude que le substrat est dépourvu de structure secondaire. La façon dont les protéines SSB et RecA peuvent agir de concert sur la même portion d'ADN n'est pas claire. Comme la protéine SSB, la protéine RecA est nécessaire en quantités stœchiométriques, ce qui suggère que son action, lors de l'assimilation de brin, implique une fixation coopérative de l'ADN pour former une structure apparentée à celle du filament.

Nous pouvons diviser en trois phases la réaction entre

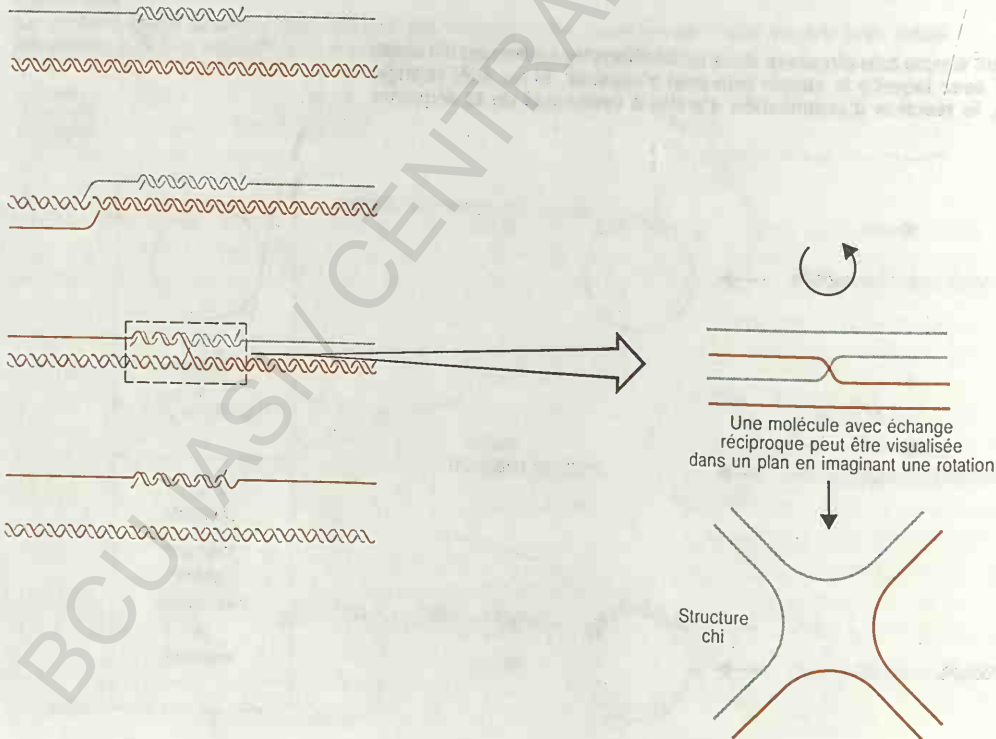
l'ADN simple brin et l'ADN double brin, catalysée par la protéine RecA :

- une phase lente présynaptique, lors de laquelle la protéine RecA polymérise sur l'ADN simple brin ;
- une réaction rapide d'appariement entre l'ADN simple brin et son complément au sein du double brin, pour former un double brin hétérologue jointif ;
- un déplacement lent d'un brin du double brin pour former une longue région d'ADN double brin hétérologue.

Lorsqu'une réaction s'établit entre une molécule simple brin et un ADN double brin, la molécule double brin se déroule dans la région de jonction avec le recombinant. La région initiale d'ADN double brin peut ne pas avoir toujours la forme parfaitement conventionnelle de la double hélice, mais pourrait être formée de deux brins associés côte à côte. Une région de ce type est appelée **jonction paranémique** (en comparaison avec la configuration **plectonémique** entrelacée des brins en double hélice).

Figure 28.10

L'échange de brin catalysé par la protéine RecA entre un ADN partiellement ou totalement double brin engendre une molécule jointive ayant la même structure qu'un intermédiaire de recombinaison.



Une jonction paranémique est instable ; la progression ultérieure de la réaction exige sa conversion en double hélice. Cette réaction revient à enlever les superenroulements négatifs et peut avoir besoin d'une enzyme résolvant le problème de déroulement/réenroulement par des cassures transitoires qui permettent aux brins de tourner l'un autour de l'autre (voir plus loin). Une conséquence de l'implication de ces enzymes est que la recombinaison peut survenir par désappariement ou appariement croisé, dans les cassures au niveau du brin initial représentées sur la Figure 28.2.

Toutes les réactions dont nous avons discutées représentent seulement une partie du phénomène potentiel de recombinaison : l'invasion d'un double brin par un simple brin. Deux molécules double brin peuvent interagir entre elles sous le contrôle de la protéine RecA, pourvu que l'une d'elles possède une région simple brin d'au moins 13 bases. La région simple brin peut prendre la forme d'une queue sur une molécule linéaire ou d'une brèche dans une molécule circulaire.

La réaction entre une molécule partiellement double brin et une molécule entièrement double brin conduit à l'échange de brins. La Figure 28.10 en illustre un exemple. L'assimilation débute à une extrémité de la molécule linéaire, à l'endroit où le brin envahisseur déplace son homologue dans le double brin selon le mode habituel. Mais lorsque la réaction atteint la région double brin dans les deux molécules, le brin envahisseur se sépare de son partenaire, qui alors s'apparie avec l'autre brin déplacé.

A ce stade, la molécule a une structure indistincte de celle de la jonction recombinante représentée sur la Figure 28.2. Lorsque les molécules qui réagissent sont circulaires, le produit est visualisé sous forme de structure *chi*. La réaction contrôlée *in vitro* par la protéine RecA peut engendrer des structures *chi*, ce qui suggère que l'enzyme peut servir d'intermédiaire à un transfert réciproque de brin.

Nous avons considéré les réactions de la protéine RecA dans un contexte indépendant de la séquence de l'ADN. Cependant, certaines zones sensibles (hot-spots) stimulent le système de recombinaison RecA. Elles ont été découvertes dans le phage lambda sous forme de mutants, appelés *chi*, qui ont des modifications portant sur une seule paire de bases, créant des sites qui stimulent la recombinaison.

Ces sites se répartissent sur la même séquence asymétrique de 8 pb :

5' GCTGGTGG 3'
3' CGACCACC 5'

La séquence *chi* se rencontre naturellement dans l'ADN d'*E. coli* environ une fois tous les 10 kb. Son absence

dans l'ADN de lambda sauvage, et également dans d'autres éléments génétiques, prouve qu'elle n'est pas essentielle à la recombinaison catalysée par la protéine RecA.

Une séquence *chi* stimule la recombinaison sur une zone étendue, c'est-à-dire dans les limites d'une distance supérieure à 10 kb du site. Un site *chi* peut être activé par une cassure d'un double brin survenue à plusieurs kilobases d'un côté précis (à droite de la séquence comme il a déjà été décrit). Cette dépendance d'orientation suggère que le système recombinant doit s'unir à l'ADN sur une extrémité cassée, puis se déplacer le long du double brin dans seulement une direction.

Les séquences *chi* peuvent identifier les sites où la protéine RecA amorce une réaction. La protéine RecA a besoin d'une région simple brin sur l'une des molécules d'ADN réagissantes. L'endonucléase RecBC peut être impliquée soit dans la dégradation d'un brin, soit dans le déroulement de l'ADN pour produire ce site *in vivo*. Identifiée à l'origine comme une nucléase puissante, l'enzyme en présence de SSB est également capable de dérouler l'ADN pour donner des régions simple brin. (Son nom révèle que ses sous-unités sont codées par les gènes *recB* et *recC*).

La dépendance de l'orientation dans la séquence *chi* se rapporte à l'influence de la nucléase RecBC. Lorsque l'enzyme se fixe sur l'ADN à droite du site *chi*, elle se déplace en déroulant l'ADN. Lorsqu'elle atteint le site *chi*, elle clive un brin d'ADN (le supérieur) en une position située entre 4 et 6 bases sur la droite. L'action de la nucléase est décrite sur la Figure 28.11 ; elle se déplace vers le site *chi*, et s'arrête pour le cliver lorsqu'elle l'atteint.

Ce modèle repose sur les données obtenues à partir de réactions *in vitro* ; nous ne connaissons pas exactement le rapport entre ces réactions et la recombinaison *in vivo*, mais il paraît probable que le déroulement et le clivage catalysés par la protéine RecBC sont impliqués dans la formation d'extrémités qui amorcent l'élaboration de la jonction de double-brins hétérologues. Il est possible que les extrémités soient captées par la protéine RecA, qui a besoin d'extrémités libres pour amorcer la réaction d'assimilation de brin.

Les activités de la protéine RecA sont-elles typiques de celles impliquées dans la recombinaison chez d'autres organismes ? Une protéine capable de manipuler l'ADN d'une manière identique à la protéine RecA a été purifiée à partir du champignon *Ustilago maydis*. La protéine Rec1 peut relier en synapse des molécules d'ADN double brin homologues avec la même dépendance vis-à-vis du simple brin que la protéine RecA. En présence d'autres enzymes, les molécules d'ADN reliées en synapse peuvent être enchaînées.

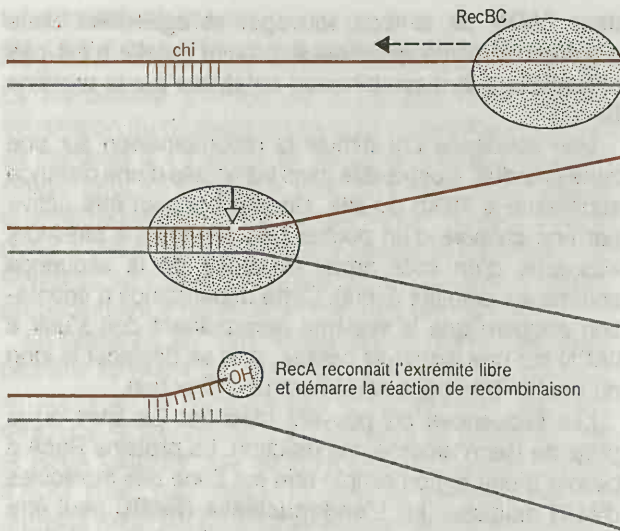


Figure 28.11
La nucléase RecBC, venant de la droite, se dirige vers une séquence *chi* et la clive de ce côté.

LA CONVERSION GÉNIQUE EST RESPONSABLE DE LA RECOMBINAISON INTERALLÉLIQUE

L'implication d'ADN double brin hétérologue dans la recombinaison explique les caractéristiques de la recombinaison entre allèles ; en effet, ces données furent à l'origine du modèle double brin hétérologue. Lorsque la recombinaison entre allèles fut découverte, la supposition naturelle fut qu'elle se produisait selon le même mécanisme de recombinaison réciproque qui s'applique à des loci plus éloignés. C'est-à-dire qu'un phénomène particulier de cassure et soudure se produit à l'intérieur du locus pour donner une paire réciproque de chromosomes recombinants. Cependant, pour ce qui est d'un gène unique donné, c'est la formation d'un double brin d'ADN hétérologue qui est elle-même responsable de l'événement de recombinaison.

Des événements particuliers de recombinaison peuvent être étudiés chez les champignons *Ascomycetes*, car les produits d'une seule méiose sont rassemblés dans une grande cellule, l'asque. Mieux même, les quatre noyaux haploïdes produits lors de la méiose sont disposés en un ordre linéaire. En fait, une mitose survient après la production de ces quatre noyaux, donnant des séries linéaires de 8 noyaux haploïdes. La **Figure 28.12** montre que chacun de ces noyaux représente effectivement le caractère génétique de l'un des huit brins des quatre chromosomes issus de la méiose.

La méiose chez un hétérozygote peut donner quatre copies de chaque allèle. C'est le cas de la majorité des spores. Mais, pour quelques spores, ces taux sont anormaux. Cela s'explique par la formation et la correction d'ADN double brin hétérologue dans la région où les allèles diffèrent.

Supposons que deux allèles diffèrent par une seule mutation ponctuelle. Lorsqu'un échange de brins survient pour former un double brin d'ADN hétérologue, les deux brins seront mal appariés au site de la mutation. En effet, chaque brin d'ADN porte une information génétique différente. Si aucun changement n'apparaît dans la séquence, les brins se séparent lors de la réplication suivante, chacun donnant naissance à un double brin qui perpétue son information. Il en résulte un rapport 4:4 anormal en ce que l'ordre des spores est altéré, en raison des échanges simple brin. Ce phénomène est appelé **ségrégation postméiotique**, parce que son existence dépend de la séparation des brins d'ADN lors de la réplication qui succède à la méiose.

Certains asques montrent des rapports de 3:5 ou 2:6 dans lesquels, respectivement, une ou deux spores, qui auraient dû appartenir à l'un des types alléliques, est, en fait, d'un autre type. Ceci provient de la correction de défauts d'appariement (mismatches) dans l'ADN double brin hétérologue.

Lorsqu'un système de réparation reconnaît des bases mal appariées dans un ADN double brin hétérologue, il peut exciser et remplacer un des brins pour restaurer la complémentarité. Un tel phénomène échange le brin d'ADN représentant un allèle contre la séquence de l'autre allèle. Si cela se produit sur un seul des doubles brins réciproques, il en résulte un rapport 3:5 ou 5:3 (selon le sens de la conversion). Si les deux doubles brins sont corrigés de la même façon, il en résulte un rapport 2:6 ou 6:2.

Le processus de correction est appelé **conversion génique** ; on le reconnaît à n'importe quel rapport aberrant. (Chaque sens de conversion est également probable, ou bien les effets d'allèle spécifique peuvent procurer une préférence pour une direction.)

La conversion génique ne dépend pas du crossing-over, mais lui est corrélée. Une grande proportion d'asques aberrants montre une recombinaison génétique entre deux marqueurs de chaque côté d'un site de conversion génique interallélique. C'est exactement ce qui aurait pu être prévu, si les rapports aberrants résultaient de la formation d'un ADN double brin, avec une probabilité approximativement égale de résoudre la structure avec ou sans recombinaison (comme le montre la **Figure 28.2**). L'implication de cela est que les chromoso-

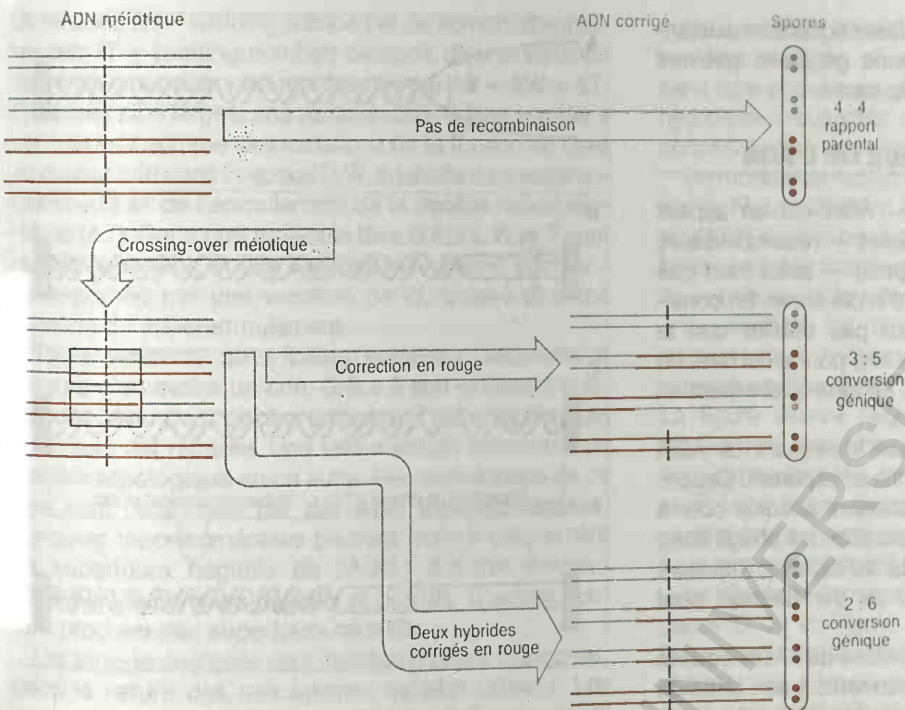


Figure 28.12 La formation de spores chez les *Ascomycètes* permet de déterminer la constitution génétique de chacun des brins d'ADN impliqués dans la méiose.

Une recombinaison provoque la formation d'ADN hybride dans la région encadrée. Le destin génétique d'un locus, dans cette région, est indiqué par une ligne verticale. Si l'une des molécules d'ADN est corrigée, les spores se séparent selon le rapport 3 : 5. Si les deux sont corrigées de la même façon, le rapport est de 2 : 6. Chaque spore représente un brin d'ADN, parce qu'une nouvelle réplication survient entre la méiose et la formation de spores.

mes des champignons initient des crossing-over environ deux fois plus souvent que la fréquence de recombinaison entre gènes éloignés pourrait le laisser supposer.

A l'intérieur d'un gène, une recombinaison entre deux sites survient lorsqu'un ADN double brin hétérologue est formé ou est corrigé au niveau d'un site mais pas de l'autre. Pourquoi ce processus permet-il l'établissement d'une carte génétique linéaire ? Plus deux sites sont rapprochés, plus grande est la probabilité qu'ils subissent une **co-conversion** sur la même région d'ADN double brin hétérologue. Dans ce cas, les deux sites sont corrigés de la même façon, inversant les proportions de génotypes issus du croisement, mais il n'en résulte pas une recombinaison entre les marqueurs.

Le double brin d'ADN hétérologue peut s'étendre sur des distances appréciables. Quelques renseignements sur l'étendue de la conversion génique sont fournis par

les séquences des membres des familles de gènes. Habituellement, les produits d'un événement de recombinaison se séparent et ne sont, par conséquent, plus disponibles pour être analysés au niveau des séquences d'ADN. Cependant, si un échange non équationnel a lieu entre deux membres d'une famille de gènes, comme l'a montré la Figure 19.3, un ADN double brin hétérologue peut se former entre les deux gènes non alléliques. Cet ADN double brin hétérologue peut subir une conversion génique convertissant l'un des gènes non alléliques en la séquence de l'autre.

La présence de plus d'une copie de gène sur le même chromosome fournit une empreinte de ces événements. Par exemple, si la formation du double brin et la conversion génique survenaient au niveau d'une partie d'un gène, cette région pourrait avoir une séquence identique ou très étroitement liée à celle de l'autre gène, la partie

conservée étant plus divergente. Des séquences existantes suggèrent que les conversions géniques peuvent s'étendre sur quelques milliers de bases.

MANIPULATION TOPOLOGIQUE DE L'ADN

La manipulation topologique de l'ADN est un aspect central de ses activités fonctionnelles — recombinaison, réplication et (peut-être) transcription — aussi bien que de l'organisation de sa structure d'ordre élevé. En considérant ces processus, il ne faut pas oublier que la structure en double brin de l'ADN est, pour ainsi dire, un obstacle qui doit être franchi par n'importe quelle réaction impliquant une séparation de brin.

Toutes les activités de synthèse impliquant un ADN double brin nécessitent que les brins se séparent. Cependant, les brins ne sont pas simplement accolés côte à côte ; ils sont entrelacés. Leur séparation les oblige donc à tourner l'un autour de l'autre dans l'espace. Plusieurs aspects possibles du processus de déroulement sont illustrés par la **Figure 28.13**.

Nous pourrions imaginer la structure de l'ADN sous la forme d'une extrémité libre qui permettrait aux brins de tourner autour de l'axe de la double hélice pour se dérouler (partie A). Cependant, étant donnée la longueur de la double hélice, cela impliquerait, lors de la séparation des brins, des battements d'une grande ampleur, ce qui paraît peu probable dans les limites de la cellule.

Un résultat identique est obtenu en utilisant un appareil qui contrôle la rotation au niveau de l'extrémité libre (partie B). Néanmoins, le résultat doit être transmis au-delà d'une distance considérable, impliquant la rotation d'une longueur de matériel démesurée.

L'ADN se comporte réellement comme une structure fermée sans extrémités libres (voir Chapitre 25), ce qui nous empêche d'établir des règles à partir de ces modèles, et nous replace face à la complexité du problème topologique, considérant les effets de séparation des deux brins dans une molécule dont les extrémités ne peuvent tourner librement (partie C). Lorsque deux brins entrelacés sont séparés à partir d'une extrémité, ceci a pour effet d'*augmenter leur enroulement l'un autour de l'autre, plus loin sur la molécule*. Ainsi, le mouvement d'une fourche de réplication pourrait engendrer un superenroulement positif accru à l'avant de celle-ci, entraînant rapidement une résistance insurmontable à tout mouvement ultérieur.

Le problème peut être surmonté en introduisant une brèche transitoire sur un brin. La partie D de la figure montre que l'extrémité libre interne permet au brin coupé de tourner autour du brin intact, après quoi la brèche

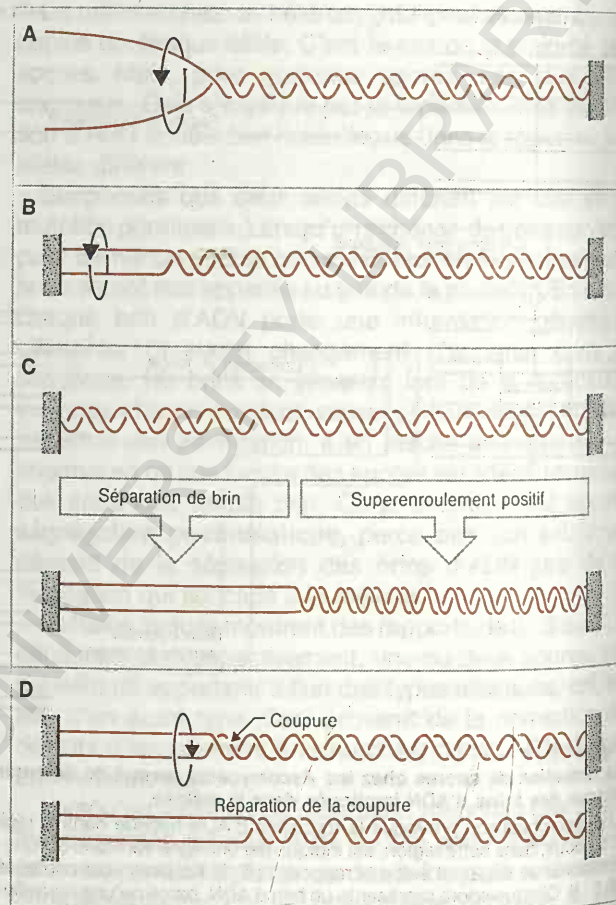


Figure 28.13

La séparation des brins d'une double hélice d'ADN peut se faire de plusieurs manières.

- A : rotation autour d'une extrémité libre.
- B : un appareil maintient les brins pendant la rotation.
- C : superenroulement positif compensatoire en un point quelconque.
- D : création d'une coupure, rotation et réparation de la coupure.

peut être soudée. La coupure et la soudure peuvent se répéter au cours de la progression de la fourche de réplication.

Nous avons vu au Chapitre 3 qu'une molécule fermée d'ADN peut être caractérisée par son nombre d'entrecroisements (linking number), c'est-à-dire le nombre de fois qu'une chaîne croise l'autre dans l'espace. Les molécules d'ADN fermées, de séquence identique, peuvent avoir des nombres différents d'entrecroisements, reflétant différents degrés de superenroulement. Les molécules d'ADN identiques, qui ne diffèrent que par leur « linking number », sont appelées **isomères topologiques**.

Le « linking number » est égal à la somme du nombre

de torsions (W = writhing number) et du nombre d'enroulements (T = twisting number) de sorte qu'une variation du « linking number » est représentée par $\Delta L = \Delta W + \Delta T$. En termes plus larges, une variation du « linking number » est égal à la somme des variations de la torsion de l'axe du double brin dans l'espace (ΔW , équivalant au superenroulement) et de l'enroulement de la double hélice elle-même (ΔT). Dans une molécule libre d'ADN, W et T sont ajustés librement, et une variation du « linking number » est exprimée par une variation de W , c'est-à-dire une variation du superenroulement.

Toute variation du « linking number » nécessite la cassure d'au moins un brin. Grâce à son extrémité libre, l'un des brins peut pivoter autour de l'autre, après quoi la cassure est réparée. Une telle réaction transforme un isomère topologique en un autre. Des conversions de ce type sont catalysées par des **ADN topo-isomérases**. Certaines topo-isomérases peuvent défaire uniquement les supertours négatifs de l'ADN ; d'autres peuvent défaire les supertours négatifs et positifs. D'autres peuvent produire des supertours négatifs.

Les topo-isomérases sont divisées en deux catégories, selon la nature des mécanismes qu'elles utilisent. Les enzymes de **type I** agissent en introduisant une cassure transitoire sur un brin d'ADN. Les enzymes de **type II** introduisent une cassure transitoire double brin. De même que les enzymes qui se comportent comme des topo-isomérases générales agissant sur l'ADN sans spécificité de séquence, les enzymes impliquées dans les recombinaisons spécifiques de sites peuvent être définies comme des topo-isomérases (voir plus loin).

La topo-isomérase de type I, la mieux connue, est le produit du gène *topA* d'*E. coli* qui relâche les contraintes dues au superenroulement négatif de l'ADN. L'enzyme n'agit pas sur l'ADN positivement superenroulé. Des mutations entraînent une augmentation du superenroulement dans le nucléoïde (et peuvent affecter la transcription, voir Chapitre 9).

En plus du relâchement des supertours négatifs dans l'ADN double brin, l'enzyme interagit avec l'ADN simple brin. Il peut préférer les supertours négatifs parce qu'ils contribuent à stabiliser les régions simple brin ; ce qui pourrait fournir le substrat sur lequel se lie l'enzyme.

Lorsque la topo-isomérase I d'*E. coli* se lie à l'ADN, elle forme un complexe stable dans lequel un brin de l'ADN a été coupé, alors que son extrémité 5' phosphate possède une liaison covalente avec un résidu tyrosine de l'enzyme. Cela suggère un mécanisme possible pour le rôle de l'enzyme ; il transfère une liaison phosphodiester de l'ADN sur la protéine, manipule la structure des deux brins d'ADN, et enfin répare le brin d'origine.

Le transfert de liaisons d'un acide nucléique à une protéine explique pourquoi l'enzyme peut fonctionner sans faire appel à un apport d'énergie. Il n'y a eu aucune hydrolyse irréversible des liaisons ; leur énergie a été conservée lors des réactions de transfert.

Un modèle de l'action de la topo-isomérase I est illustré sur la **Figure 28.14**. L'enzyme est fixée sur une région où l'ADN double brin commence à se séparer en simples brins ; puis elle casse un brin, passe l'autre brin au travers de la brèche, et la referme.

La réaction fait changer le « linking number » par fractions de 1. A chaque fois qu'un brin passe à travers la brèche créée dans l'autre brin, il y a un ΔL égal à + 1. La figure illustre l'activité de l'enzyme en terme de mouvement des brins individuels ; dans une molécule libre superenroulée, le caractère interchangeable de W et T conduit un changement du « linking number » à correspondre à une variation de $\Delta W = + 1$, c'est-à-dire à un supertour négatif en moins (voir Chapitre 3).

La réaction correspond à la rotation décrite dans la partie *D* de la Figure 28.13, à la différence près que l'enzyme limite la réaction à un seul passage de brin à chaque fois. (Au contraire, l'introduction d'une brèche dans une molécule superenroulée permet d'alléger la tension par de nombreuses rotations du brin libre lors de chaque réaction.)

La topo-isomérase I peut faire passer un fragment d'un simple brin d'ADN au travers d'un autre. Cette réaction de **passage simple brin** peut introduire des **nœuds** dans l'ADN et peut **unir** deux molécules circulaires comme des maillons d'une chaîne. Nous ne savons pas encore à quels usages (s'il y en a) ces réactions sont appliquées *in vivo*.

Les propriétés formelles des topo-isomérases de type I chez les eucaryotes sont identiques, mais elles peuvent relâcher les supertours positifs et négatifs. Leur mode d'action est différent : au lieu d'utiliser des passages simples brins, dans lesquels l'une des extrémités de la cassure est entravée, elles se comportent comme des tourniquets, c'est-à-dire qu'elles coupent un brin puis ferment la brèche après une rotation.

Les topo-isomérases de type II relâchent en général les supertours négatifs et positifs. La réaction nécessite la présence d'ATP ; probablement, un ATP est hydrolysé lors de chaque réaction catalytique. Comme le montre la **Figure 28.15**, la réaction se fait par une cassure double brin dans un ADN double brin, et le passage au travers de celui-ci, d'une autre région double brin.

La conséquence qui découle d'un transfert de deux brins est que le « linking number » varie par multiples de 2. L'activité de la topo-isomérase II peut servir aussi à

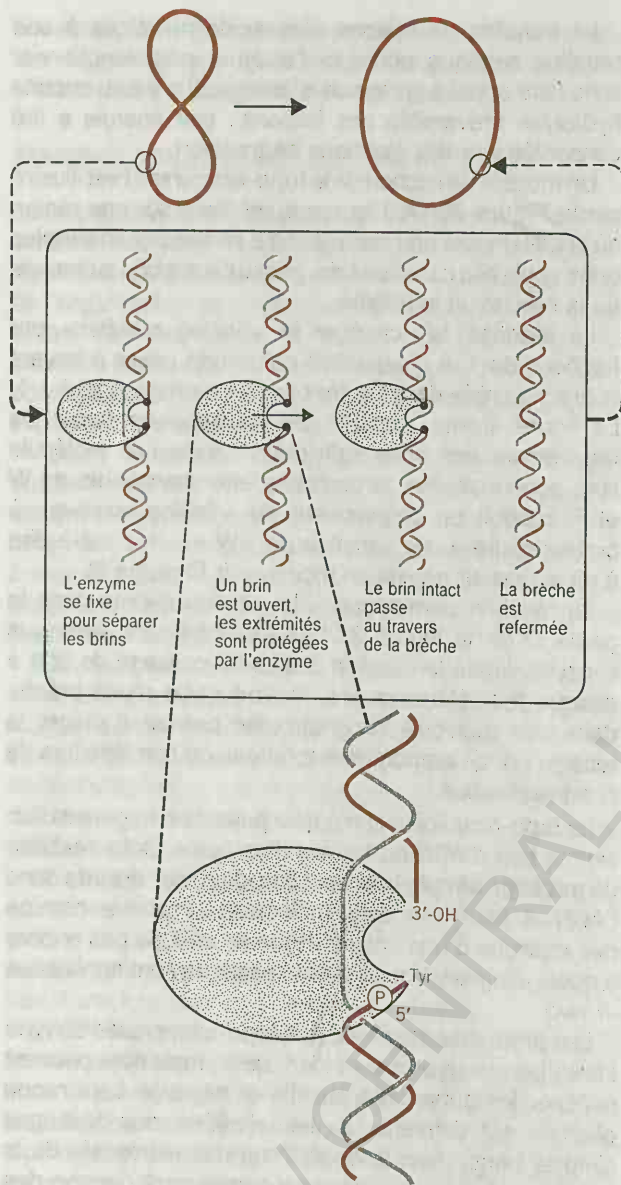


Figure 28.14

Les topo-isomérases bactériennes de type I reconnaissent partiellement les fragments déroulés d'ADN et font passer un brin à travers une cassure faite dans l'autre.

faire et à défaire des double brins circulaires enchaînés et des molécules nouées.

La réaction représente probablement une reconnaissance non spécifique de l'ADN double brin sur lequel l'enzyme se lie à n'importe quels fragments double brin qui s'entrecroisent. L'hydrolyse de l'ATP peut servir à faire subir à l'enzyme des changements de conformation qui fournissent la force nécessaire pour pousser un ADN double brin à travers la brèche faite dans l'autre double brin. En raison de la topologie de l'ADN superenroulé, le croisement des segments d'ADN double brin permet de retirer des supertours d'un cercle, qu'il soit positivement ou négativement superenroulé. Le meilleur exemple connu de ce type d'enzyme est peut-être la topo-isomérase II de T4, qui est le produit de 3 gènes, et dont l'activité est essentielle à la réplication du phage.

LA GYRASE INTRODUIT DES SUPERTOIRS NÉGATIFS DANS L'ADN

L'ADN gyrase bactérienne est une topo-isomérase de type II qui est capable d'*introduire* des supertours négatifs dans une molécule circulaire fermée relâchée. L'ADN gyrase se fixe sur un ADN circulaire double brin et catalyse un superenroulement progressif : elle continue à introduire des supertours dans la même molécule d'ADN. Une molécule d'ADN gyrase peut introduire environ 100 supertours par minute.

La forme superenroulée de l'ADN a une énergie libre plus importante que la forme relâchée, et l'énergie nécessaire à la conversion est fournie par l'hydrolyse de l'ATP. En l'absence d'ATP, la gyrase peut *relâcher* des supertours négatifs mais pas des supertours positifs, bien que la vitesse soit 10 fois inférieure à la vitesse d'introduction des supertours.

La structure de l'ADN gyrase d'*E. coli* est résumée dans le **Tableau 28.1**. L'enzyme est inhibée par deux types d'antibiotique, chacun agissant sur l'une des sous-unités. Les médicaments inhibent la réplication, ce qui suggère que l'ADN gyrase est nécessaire à la synthèse de l'ADN. Les mutations qui confèrent une résistance aux antibiotiques caractérisent les loci codant pour les sous-unités.

La gyrase fixe son ADN substrat autour de la face externe de la protéine tétramérique. La gyrase protège de la digestion par la nucléase du micrococcoque, environ 140 pb d'ADN (protection très proche de celle fournie par l'octamère histone beaucoup plus petit).

Le modèle d'*inversion de signe* caractérisant l'action de la gyrase est illustré par la **Figure 28.16**. L'enzyme se

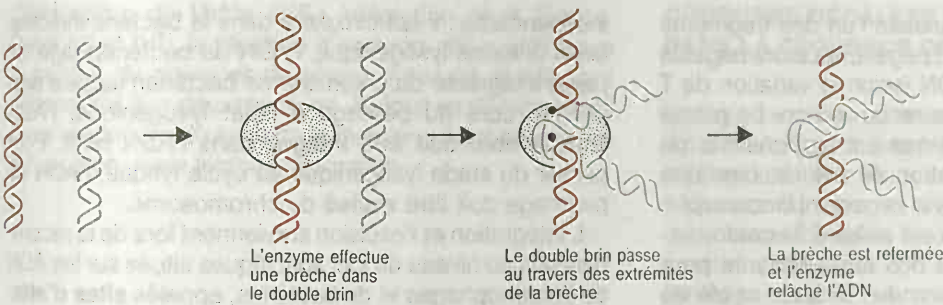


Figure 28.15

La topo-isomérase type II peut faire passer un ADN double brin à travers une cassure double brin produite dans un autre ADN.

Tableau 28.1

La gyrase comporte deux sous-unités servant de cible à différents antibiotiques.

Sous-unité	Taille	Locus	Antibiotiques agissant sur la sous-unité
A	105 000 daltons	<i>gyrA</i> (<i>nalA</i>)	acide nalidixique et acide oxilinique
B	95 000 daltons	<i>gyrB</i> (<i>cou</i>)	coumermycine A1 et novobiocine

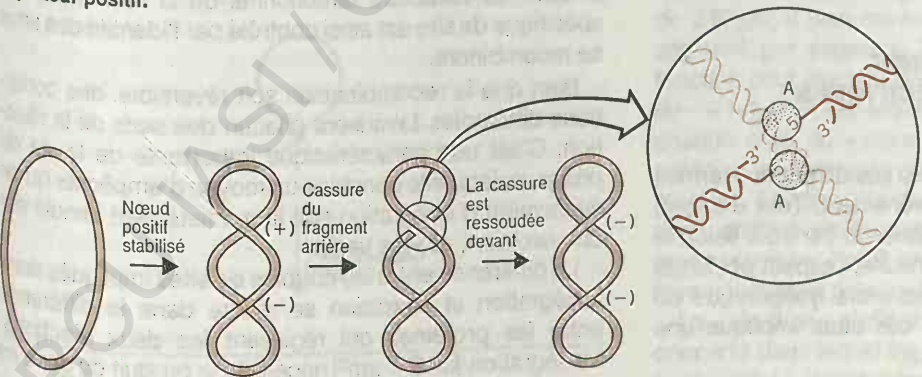
La gyrase est un tétramère de 400 000 daltons de structure A_2B_2 .

fixe à l'ADN dans une configuration croisée équivalente à un supertour positif. Ceci induit un supertour négatif compensatoire dans l'ADN libre. Ensuite, l'enzyme produit une cassure du double brin au niveau du croisement de supertour positif et fait passer l'autre double brin au travers de la brèche qu'elle referme ensuite.

La réaction inverse le signe du supertour : il a été converti de + 1 tour à - 1 tour. Ainsi le « linking number » a varié de $\Delta L = - 2$, en accord avec l'exigence que tous les événements impliquant un passage du double brin doivent faire varier le « linking number » d'un multiple de deux.

Figure 28.16

L'ADN gyrase peut introduire des supertours négatifs dans un double brin d'ADN en inversant un supertour positif.



Lorsque la gyrase relâche ensuite l'un des fragments croisés du supertour lié devenu négatif, les tours négatifs se répartissent le long de l'ADN (selon la variation de T ou W ou des deux) et le cycle recommence. Le même type de manipulation topologique est responsable de l'enchaînement et de la formation de nœuds, bien que les rôles de ces réarrangements *in vivo* soient inconnus.

Lorsqu'un supertour inversé est relâché, la conformation de la gyrase change. Elle doit être restaurée pour permettre à l'enzyme d'entreprendre un autre cycle de superenroulement. Ce processus s'appelle le **renouvellement de l'enzyme**. On suppose qu'il est entraîné par l'hydrolyse de l'ATP, puisque le remplacement de l'ATP par un analogue non hydrolysable permet à la gyrase de n'introduire qu'une inversion (moins 2 supertours) par substrat. Ainsi, l'ATP n'est pas nécessaire à la réaction de superenroulement, mais est indispensable pour qu'elle entreprenne un second cycle. La novobiocine interfère avec les réactions ATP-dépendantes de la gyrase, en empêchant l'ATP de se lier à la sous-unité B.

La réaction de relâchement, indépendante de l'ATP, est inhibée par l'acide nalidixique. Ceci permet d'impliquer la sous-unité A dans le processus de coupure et de soudure. Le traitement de la gyrase par l'acide nalidixique permet d'aboutir à des fragments d'ADN qui sont engendrés par des clivages décalés, des deux brins de la molécule d'ADN. Les extrémités possèdent toutes un groupement hydroxyle (OH) libre en position 3', et une extension de 4 bases en 5', unie par une liaison covalente à la sous-unité A (voir Figure 28.16). La liaison covalente conserve l'énergie de la liaison phosphate ; cette énergie peut être utilisée pour mener la réaction de la soudure, expliquant pourquoi la gyrase peut entreprendre un relâchement sans ATP. Les sites de clivage sont tout à fait spécifiques, on les rencontre environ une fois toutes les 100 pb.

UNE RECOMBINAISON DONNÉE RECONNAÎT DES SITES SPÉCIFIQUES

La conversion de l'ADN lambda en ses différentes formes de vie implique deux types d'événement. Tout d'abord, le schéma de l'expression du gène est contrôlé selon le modèle décrit au chapitre 12. Ensuite, l'aspect physique de l'ADN est différent, selon les états lysogénique ou lytique ; l'interconversion entre ces états implique une recombinaison spécifique de site.

Sous la forme lytique, l'ADN lambda est une molécule

indépendante ; il est circulaire dans la bactérie infectée. Sous la forme lysogénique, l'ADN du bactériophage fait partie intégrante du chromosome bactérien (appelé prophage). Lors du passage à l'état lysogénique, l'ADN lambda libre doit être **intégré** dans l'ADN hôte. Pour passer du stade lysogénique au cycle lytique, l'ADN du prophage doit être **excisé** du chromosome.

L'intégration et l'excision surviennent lors de la recombinaison au niveau de loci spécifiques situés sur les ADN de bactériophages et de bactéries, appelés **sites d'attachement (att)**. Le site d'attachement sur le chromosome bactérien est appelé *att^B* en génétique bactérienne. Le locus est caractérisé par des mutations qui empêchent l'intégration du lambda ; il est occupé par le prophage λ dans les souches lysogéniques.

Pour analyser la recombinaison illustrée par la Figure 28.17, nous appellerons le site de fixation bactérien *att^B*, constitué par les séquences *BOB'*. Le site de fixation du phage, *att^P*, est constitué par les composants *POP'*.

La séquence O est commune à *att^B* et *att^P*. On l'appelle le « cœur » de la séquence ; la recombinaison a lieu à l'intérieur de cette séquence. Les régions bordantes B, B' et P, P' sont considérées comme les **bras** ; chacun a une séquence distincte.

L'ADN de phage étant circulaire, il est inséré lors de la recombinaison, comme une séquence linéaire dans le chromosome bactérien. Le prophage est fixé sur deux nouveaux sites *att*, produits de la recombinaison. Selon la représentation habituelle de la carte, on trouve à gauche le site *att^L* constitué par *BOP'*, et à droite le site *att^R* constitué par *POB'*.

Une conséquence importante de la constitution des sites *att* est que l'intégration et l'excision n'impliquent pas la même paire de séquences. L'intégration exige une reconnaissance entre les sites *att^P* et *att^B* ; tandis que l'excision exige une reconnaissance entre les sites *att^L* et *att^R*. Le caractère directionnel de la recombinaison spécifique de site est ainsi contrôlé par l'identité des sites se recombinant.

Bien que la recombinaison soit réversible, des conditions différentes favorisent chacun des sens de la réaction. C'est une caractéristique importante de la vie du phage, puisqu'elle constitue un moyen d'empêcher qu'un événement d'intégration soit immédiatement annulé par une excision, et vice versa.

La différence entre les couples de sites impliqués dans l'intégration et l'excision se reflète dans la différence entre les protéines qui régissent les deux réactions. L'intégration (*att^B × att^P*) nécessite le produit du gène *int* du phage et une protéine bactérienne appelée facteur

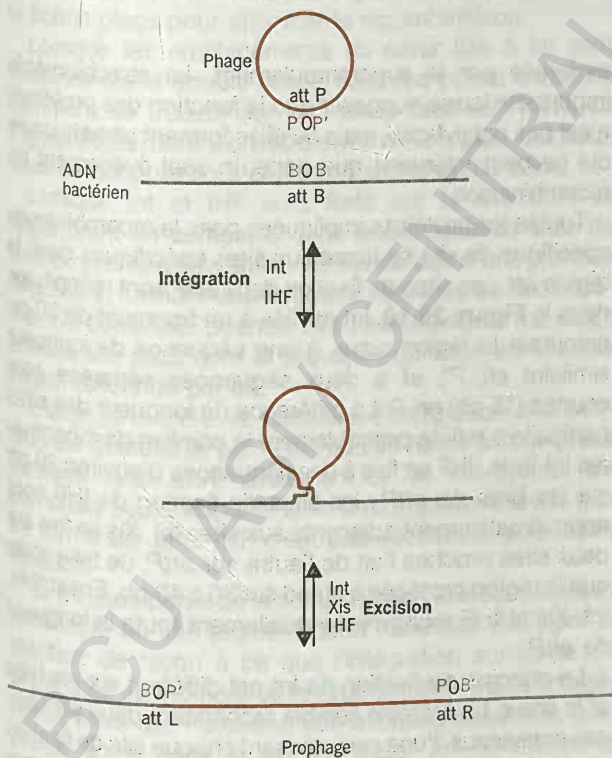
d'intégration de l'hôte (IHF : *Intégration Host Factor*). Outre *Int* et IHF, la réaction d'excision *attL* $\times$ *attR* exige le produit du gène *xis* du phage. Ainsi *Int* et IHF sont nécessaires aux deux réactions. *Xis* joue un rôle important dans le contrôle du sens de la réaction ; il est nécessaire à l'excision, mais inhibe l'intégration.

Figure 28.17

Le phage d'ADN circulaire est transformé en prophage intégré par une recombinaison réciproque entre les sites *attP* et *attB* ; le prophage est excisé par une recombinaison réciproque entre les sites *attL* et *attR*.

Lorsque le site *att^λ* est délété sur le chromosome d'*E. coli*, un bactériophage lambda infectant peut induire une lysogénicité par une intégration dans un autre site, bien que l'efficacité de la réaction représente moins de 0,1 % de la fréquence d'intégration au niveau du site *att^λ*. Cette intégration inefficace se produit au niveau des sites d'attachement secondaires, qui ressemblent à d'authentiques séquence *att*.

Un aspect important de ces réactions secondaires est l'homologie dans le chevauchement de 7 pb entre les couples de sites recombinants ; un autre aspect est représenté par l'homologie avec les séquences reconnues par *Int* autour des sites de coupure.



COUPURES DÉCALÉES ET RESSOUDURE DANS LA SÉQUENCE CŒUR

La région « cœur » se trouve dans une séquence de 15 pb riche en A-T commune à tous les sites *att*. Les recombinaisons ont-elles lieu en n'importe quelle position de la séquence du cœur ou en un point spécifique ?

La recombinaison spécifique de site implique une cassure précise, et une soudure en l'absence de toute synthèse d'ADN. Les points d'échange sont différents sur chaque brin d'ADN.

Le modèle illustré par la Figure 28.18 montre que si les sites *attP* et *attB* ont subi chacun les mêmes coupures décalées, les extrémités simple brin complémentaires peuvent être disponibles pour l'hybridation croisée. La distance entre les points de croisement du lambda est de 7 pb, et la réaction engendre des extrémités 3'-phosphate et 5'-OH.

Une mutation survenant de chaque côté de la zone de recouvrement de 7 pb est condamnée à demeurer sur le bras où elle se trouve. Elle ne peut ainsi jamais être transférée d'un site *attB* sur un site *attP* (ou d'un site *attL* sur un site *attR*). Une mutation dans la zone de 7 pb peut cependant être transférée par une conversion génique entre les sites *attP* et *attB* (et peut ainsi se déplacer entre tous les sites *att*). La découverte d'un tel transfert a servi de fondement à l'idée originale selon laquelle tous les sites *att* doivent avoir un « cœur » commun, bien que la séquence habituelle (15 pb) se trouve être plus longue que la région impliquée dans le décalage entre les coupures des brins.

La recombinaison spécifique de site peut être effectuée *in vitro* grâce à *Int* et IHF. Le rôle des bras peut être exploré en produisant des délétions de chaque côté. La fonction du site *attP* est abolie par des délétions qui réduisent sa taille au-delà des limites de -152 en P à +82 en P'. Pour déterminer le rôle de régions individuelles dans cette zone de 240 pb, il sera nécessaire d'introduire des variations segment par segment. Le site *attB* est plus petit ; sa fonction peut être exercée par un fragment de 23 pb allant de -11 à +11, sur lequel se trouvent seulement 4 pb de chaque côté du « cœur ». Un fragment d'*attP* de taille similaire se comporte comme un site *attB* lors de cette réaction. La disparité dans leurs tailles suggère que les sites *attP* et *attB* jouent des rôles différents dans la recombinaison, l'*attP* donnant l'information supplémentaire nécessaire pour le distinguer de *attB*.

La réaction s'effectue-t-elle par un mécanisme concerté dans lequel les brins dans les sites *attP* et *attB* sont coupés simultanément et échangés ? Ou bien les

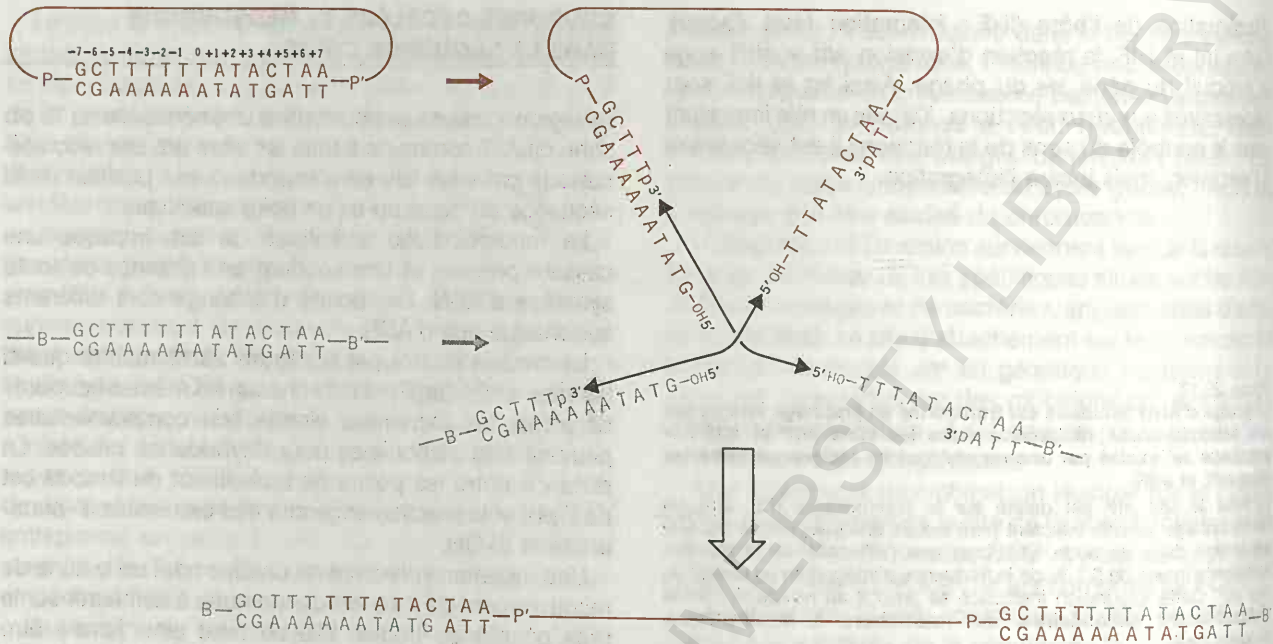


Figure 28.18

Les clivages décalés au niveau de la séquence « cœur » commune aux sites d'*attP* et d'*attB* permettent une jonction croisée, pour engendrer des liaisons recombinantes réciproques. La séquence est numérotée par rapport au centre du « cœur ».

brins sont-ils échangés l'un après l'autre, le premier échange engendrant une structure *chi*, le deuxième cycle de coupure et réparation amenant un relâchement de la structure ? La protéine Int peut résoudre la structure *chi*, ce qui suggère que cette voie est suivie, Int étant responsable des réactions de coupure et de ligation.

La réaction *in vitro* nécessite un superenroulement au niveau de l'*attP*, mais pas de l'*attB*. Lorsque la réaction se déroule *in vitro* entre deux molécules d'ADN superenroulées, la majeure partie du superenroulement est conservée par les produits. Ainsi, il ne peut y avoir d'intermédiaires libres dans lesquels une rotation de brin pourrait survenir. Cela concorde avec l'idée que la réaction passe par la formation d'une structure *chi*. La réaction de cassure et soudure peut ressembler à une activité topo-isomérase I, sauf que le brin entaillé est ressoudé à un brin entaillé n'appartenant pas au même double brin, mais à un double brin différent. [Int, en effet, a une certaine activité topo-isomérase (à dire vrai peu efficace) capable de relâcher de l'ADN superenroulé négativement.]

Des quantités importantes de protéines Int et IHF sont nécessaires à la recombinaison *in vitro*. Int et IHF se lient coopérativement sur l'*attP*, et leur affinité pour le site est

renforcée par le superenroulement. La stœchiométrie importante laisse supposer que la fonction des protéines n'est pas catalytique, mais qu'elles forment une structure qui ne peut intervenir que dans un seul événement de recombinaison.

Toutes les protéines impliquées dans la recombinaison spécifique de site, se fixent aux sites spécifiques dans la région *att*. Les sites de fixation dans *attP* sont récapitulés dans la Figure 28.19. Int est liée à un fragment de 30 pb entourant la région cœur, à une séquence de longueur similaire en *P'*, et à deux séquences séparées plus courtes (15 pb) en *P*. La différence de longueur des sites particuliers reflète probablement le nombre de monomères Int fixés. IHF se fixe à des séquences d'environ 20 pb sur les bras de *attP*; les sites de fixation de IHF sont approximativement adjacents aux sites Int. Xis se fixe sur deux sites proches l'un de l'autre sur *attP*, de telle sorte que la région protégée s'étend sur 30 à 40 pb. Ensemble, Int, Xis et IHF recouvrent virtuellement toute la longueur de *attP*.

Le principe de fixation de Int est différent sur les bras et le cœur. La protéine semble reconnaître deux séquences consensus, l'une caractérisant chaque site de fixation sur les bras, l'autre présente dans le cœur. Le consensus

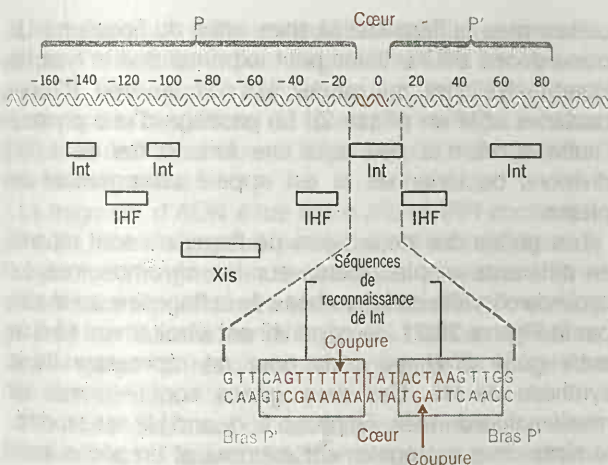


Figure 28.19
Les protéines Int et IHF se fixent sur des sites différents d'*attP*. Les séquences de reconnaissance de Int dans la région « cœur » englobent les sites de coupure.

cœur, CAACTTNNT, est trouvé avec une orientation inverse à la jonction cœur-bras des sites *attP* et *attB*. Ces derniers sites entourent les sites de coupure dans le cœur, comme le montre la Figure 28.19, situant Int à la bonne place pour effectuer la recombinaison.

Lorsque les emplacements du cœur liés à Int sont dessinés sur la double hélice, tous ses points d'attache semblent se trouver sur une seule face de l'ADN. La Figure 28.20 montre un modèle de cette région. Les deux sites de coupure sont exposés dans le grand sillon.

Lorsque Int et IHF sont fixés sur le site *attP*, ils engendrent un complexe dans lequel tous les sites de fixation peuvent être attirés vers la surface d'une protéine oligomère, formant une structure analogue au nucléosome. Ce complexe pourrait être l'intermédiaire qui réagit avec le site *attB*, dont la séquence du cœur est la seule région reconnue par Int.

L'asymétrie entre les réactions d'intégration et d'excision est démontrée par le fait que Int forme un complexe similaire avec *attR*, seulement si Xis est ajouté. Ce complexe peut s'apparier avec un complexe resserré que Int forme sur *attL*. IHF n'est pas nécessaire pour cette réaction.

La majeure partie de la complexité de la recombinaison spécifique de site provient de la nécessité de régler la réaction de façon à ce que l'intégration survienne de préférence au moment où le virus atteint le stade lysogénique, alors que l'excision est favorisée lorsque le prophage entre dans le cycle lytique. En contrôlant la quantité disponible de Int et Xis, c'est la réaction appro-

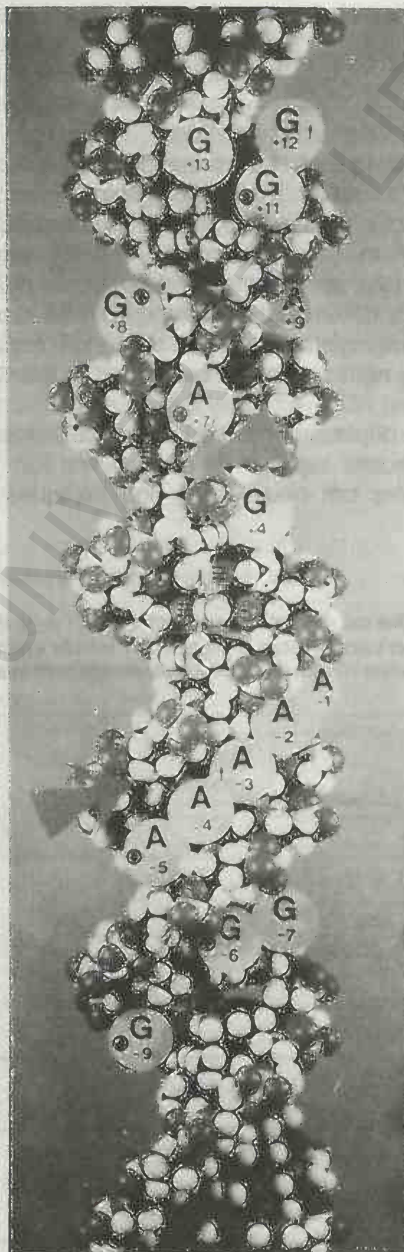


Figure 28.20
Les sites de fixation de Int dans la région « cœur » se trouvent sur une face de l'ADN.

Les cercles les plus grands indiquent les positions auxquelles la méthylation est influencée par la fixation de Int, les grosses flèches indiquent les sites de coupure.

Photographie aimablement fournie par A. Landy.

prisée qui surviendra sans risquer d'être instantanément inversée.

L'INVERSION PEUT CONTRÔLER L'EXPRESSION D'UN GÈNE

L'inversion de fragments d'ADN par recombinaison spécifique de site sert à contrôler l'expression génique dans certaines circonstances. Lors de la synthèse de la flagelline par *S. typhimurium*, une régulation +/- est fournie par le déplacement du promoteur par rapport à l'unité de transcription. Dans la détermination de la spécificité du bactériophage Mu, des gènes alternatifs sont exprimés par inversion du fragment codant par rapport au promoteur.

Les bactéries se déplacent par ondulation de leurs flagelles. De nombreuses espèces de *Salmonelles* sont **diphasiques** car elles ont deux gènes non alléliques

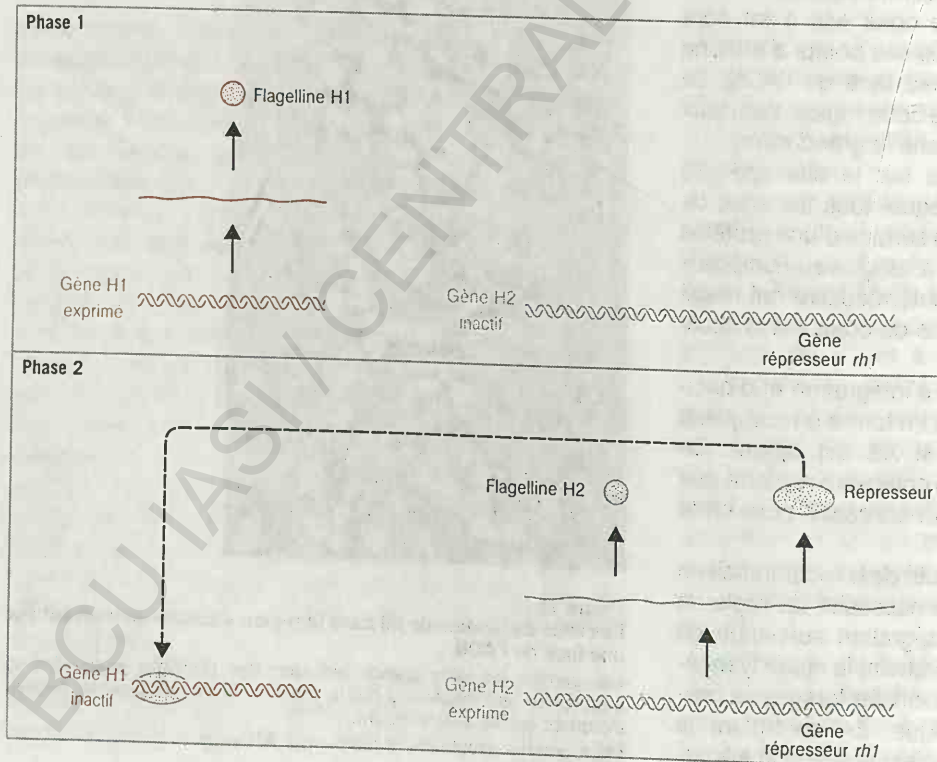
codant pour la flagelline (la sous-unité du flagellum). Un clone donné de bactéries peut exprimer soit le type H1 (c'est-à-dire être en **phase 1**), soit le type H2 (les bactéries sont en **phase 2**). Le passage d'une phase à l'autre survient à peu près une fois toutes les 1 000 divisions bactériennes et est appelé **changement de phase**.

Les gènes des deux types de flagelline sont répartis en différents emplacements sur les chromosomes. Le cycle de contrôle de la synthèse de la flagelline est illustré par la **Figure 28.21**. Le gène H2 est étroitement lié à un autre gène (*rh1*) qui code pour un répresseur de la synthèse de H1. Ces deux gènes sont exprimés de manière coordonnée. En phase 2, quand H2 est exprimé, le répresseur est également exprimé et empêche toute synthèse de H1. En phase 1, lorsque H2 n'est pas exprimé, le répresseur ne l'est pas non plus, et la synthèse de H1 a lieu. De cette façon, la phase de la bactérie est déterminée selon que l'unité de transcription *H2-rh1* est active ou non.

Figure 28.21

La phase des *Salmonelles* est déterminée par l'activité de l'unité de transcription H2.

En phase 1, l'unité H2 est inactive, ainsi la flagelline H1 est-elle synthétisée. En phase 2, l'unité H2 est active et synthétise le flagelline H2 plus un répresseur qui empêche l'expression du gène de la flagelline H1.



L'expression de cette unité de transcription est contrôlée par l'orientation d'un fragment d'ADN qui lui est adjacent. Le fragment a une longueur de 995 pb et est bordé par des répétitions de 14 pb (*IRL* et *IRR*). Le codon d'initiation de *H2* se trouve à 16 pb à droite de la répétition inversée adjacente.

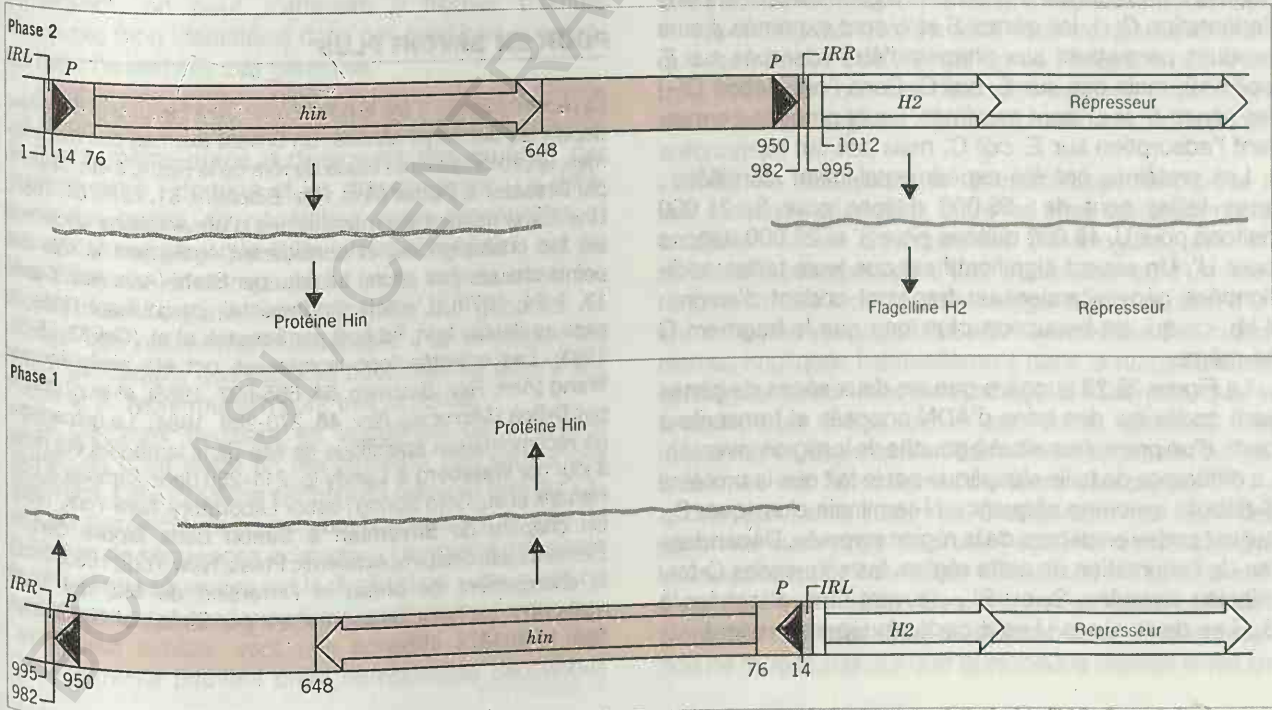
Le fragment d'ADN situé entre *IRL* et *IRR* contient le gène *hin*, dont le produit protéique est le médiateur de l'inversion du fragment entier par recombinaison réciproque entre les répétitions inversées (la réaction d'inversion générale est illustrée par la Figure 29.10). Les mutations dans le gène *hin* réduisent la fréquence de l'inversion d'environ 10⁴ fois.

Les conséquences de la réaction d'inversion sont illustrées par la Figure 28.22. Le promoteur de l'unité de transcription *H2-rh1* se trouve à l'intérieur du fragment réversible. Dans une orientation donnée, la transcription s'amorce au niveau du promoteur et se poursuit dans

H2-rh1, aboutissant à l'expression de la phase 2. Dans l'orientation opposée, le promoteur est orienté dans l'autre direction, et l'unité de transcription n'est pas exprimée (bien que la transcription ait probablement lieu en sens inverse, avec des conséquences inconnues). L'absence de transcription de *H2-rh1* aboutit à l'expression de la phase 1.

Un système comparable dans le phage Mu est révélé par un phénomène inhabituel à proximité de l'extrémité droite du génome. Le **fragment réversible** ou **fragment G**, renferme une région de 3 kb que l'on trouve orientée différemment dans différentes molécules d'ADN Mu. Dans les bactériophages cultivés par infection lytique d'*E. coli* K12, le fragment G est toujours orienté selon le sens G(+). Cependant, l'ADN de phages engendrés par induction, peut renfermer le fragment G orienté dans ce sens ou dans le sens inverse G(-). Un gène Mu appelé *gin*, situé juste au-delà du fragment G, est nécessaire à la

Figure 28.22
L'expression du groupe flagelline H2-répresseur de la phase 1 est contrôlée par l'orientation d'une région réversible de 995 pb.
La phase 2 survient lorsque le promoteur, à droite, est orienté de façon à transcrire l'unité H2 en passant par *IRR*.
La phase 1 se déroule lorsque le promoteur est orienté en sens inverse. Dans les deux phases, *hin* est exprimé indépendamment à partir de son propre promoteur.



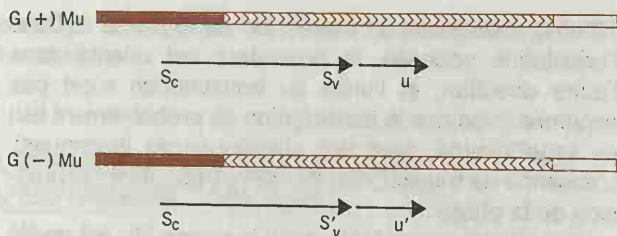


Figure 28.23

L'inversion du fragment G détermine la synthèse, soit des protéines S et U, soit des protéines S' et U'. Un modèle suppose que leur expression est amorcée à gauche du fragment réversible. Les protéines S et S' ont en commun une séquence N-terminal (S_c) codée à gauche du fragment réversible. S_v et U sont codés par un brin d'ADN tandis que S'_v et U' le sont par le brin complémentaire.

réaction d'inversion. (Une fonction d'inversion voisine, le gène *cin*, est responsable de l'inversion du fragment C du phage P1.)

Le fragment G renferme des gènes codant pour des protéines impliquées dans l'adsorption du phage. L'orientation du fragment contrôle l'expression de ces gènes, les phages G(+) et G(-) ayant des affinités différentes pour les souches de bactérie.

Comme le montre le diagramme de la Figure 28.23, le fragment G comporte des séries de gènes alternés. Dans l'orientation G(+), les gènes S et U sont exprimés. Leurs produits permettent aux phages d'être adsorbés sur *E. coli* K12, mais pas sur *E. coli* C. Dans l'orientation G(-) les gènes S' et U' sont exprimés. Leurs produits permettent l'adsorption sur *E. coli* C, mais pas sur K.

Les protéines ont été expérimentalement identifiées ; leurs tailles sont de : 56 000 daltons pour S, 21 000 daltons pour U, 48 000 daltons pour S' et 26 000 daltons pour U'. Un aspect significatif est que leurs tailles additionnées nécessiteraient un fragment codant d'environ 4 kb, ce qui est beaucoup plus long que le fragment G lui-même.

La Figure 28.23 suggère que les deux séries de gènes sont codés sur des brins d'ADN opposés et transcrits à partir d'un promoteur situé à gauche de la région inversée. La différence de taille s'explique par le fait que la protéine S débute avec une séquence N-terminale classique, S_c , qui est codée en dehors de la région inversée. Dépendantes de l'orientation de cette région, les séquences C-terminales variables, S_v ou S'_v , peuvent être rattachées à S_c . Les deux gènes U sont codés indépendamment.

La fonction *hin* des Salmonelles et la fonction *gin* de Mu peuvent se substituer l'une à l'autre lors de réactions de complémentation. Un gène voisin est également trouvé dans *E. coli*. Appelé *pin*, il est capable de compléter les mutations *gin* dans le phage Mu. Il semble catalyser l'inversion d'un fragment adjacent de 1 800 pb entouré de répétitions inversées de 29 pb. Nous ignorons le rôle de cette réaction.

Les réactions d'inversion catalysées par Gin - et Hin - peuvent être exécutées *in vitro*, où la réaction a besoin également d'une protéine codée par l'hôte, non identifiée. Le substrat doit être superenroulé. Les réactions Hin et Gin ont besoin toutes deux d'une autre séquence, ainsi que des sites de recombinaison. La séquence additionnelle est d'environ 60 pb, et peut se trouver n'importe où sur la molécule du substrat, sauf très près des sites de recombinaison eux-mêmes.

L'indépendance d'emplacement de cette séquence ressemble aux séquences stimulatrices de la transcription, ou « enhancers » (voir Chapitre 9). Les conceptions actuelles convergent vers l'idée que le « enhancer » de recombinaison sert en quelque sorte à amorcer la réaction, peut-être en fournissant un « site d'entrée » au facteur codé par l'hôte ou Hin/Gin, ou en formant probablement une structure tertiaire autour d'un des composants de la protéine.

POUR EN SAVOIR PLUS

La recombinaison a été le point commun à beaucoup de revues depuis les dernières années. Un brillant examen des mécanismes, approfondissant l'étude du rôle de la RecA, a été rapporté par Dressler & Potter (*Ann. Rev. Biochem.* 51, 727-761, 1982). Une revue relatant les mécanismes d'observations anciennes sur les champignons, et considérant également le rôle des points chauds (hot spots) fut faite par Stahl (*Ann. Rev. Genet.* 13, 7-24, 1979). L'article fondamental, introduisant l'idée de cassure double brin, fut écrit par Szostak et al. (*Cell* 33, 25-35, 1983). Les activités topo-isomérases ont été analysées par Wang (*Ann. Rev. Biochem.* 54, 665-697, 1985), et leur biologie par Drlica (*Microbiol. Rev.* 48, 273-289, 1984). Le mécanisme de recombinaison spécifique de site dans lambda a été remis à jour par Weisberg & Landy (p. 211-250 dans *Lambda II*, Eds. Hendrix et al., Cold Spring Harbor Laboratory, New York, 1983). Un chapitre de Silverman & Simon dans *Mobile Genetic Elements* (ed. Shapiro, Academic Press, New York, 1983) étudie le changement de phase et l'inversion de Mu, qui furent également explorés précédemment par Simon et al. (*Science* 209, 1370-1374, 1980).

CHAPITRE 29

LES ÉLÉMENTS TRANSPOSABLES CHEZ LES BACTÉRIES

Les génomes sont habituellement considérés comme étant un peu statiques, ne changeant que sur une longue période de temps. Nous sommes habitués à l'idée que l'élaboration d'une carte génétique permet d'identifier les loci dans lesquels se trouvent des gènes connus ; par conséquent, on peut s'attendre à trouver d'autres séquences (non identifiées) dans des positions constantes dans l'ensemble des génomes.

La stabilité de l'organisation génétique est caractérisée par le maintien de relations de liaison entre des groupes de gènes, même après la divergence des espèces, par exemple entre l'Homme et les grands singes. La différence des temps de génération entre procaryotes et eucaryotes suggère que leurs cycles d'évolution ont pu être différents en temps réel ; mais même chez les procaryotes, l'organisation générale du génome ne change que relativement lentement. Par exemple, les cartes génétiques des espèces bactériennes différentes *E. coli* et *S. typhi murium* sont très proches.

Les génomes évoluent en acquérant de nouvelles séquences mais aussi en réarrangeant des séquences existantes. Les nouvelles séquences peuvent être apportées par des vecteurs ou peuvent apparaître à la suite de mutations de séquences existantes. La source principale de nouvelles séquences est la duplication : une copie de la séquence peut garder sa fonction d'origine tandis que l'autre peut évoluer vers une nouvelle fonction. Des réarrangements peuvent créer de nouvelles séquences

et changer les fonctions de séquences existantes en les plaçant dans de nouvelles positions régulatrices.

Chez les bactéries, des éléments extrachromosomiques transmettent des informations de façon horizontale, permettant le transfert de fragments de matériel génétique (habituellement plutôt petits). Des plasmides permettent la conjugaison (croisement) entre des bactéries alors que les phages sont libérés après l'infection. Ce n'est qu'à l'occasion que ces deux phénomènes aboutissent au transfert de gènes de l'hôte, associés au réplicon autonome.

Chez les eucaryotes, la recombinaison réciproque a lieu à chaque génération entre les sites correspondants localisés sur les chromosomes homologues ; elle est rarement accompagnée d'une duplication ou d'un réarrangement des loci. Une telle réorganisation au sein d'un chromosome est surtout un effet secondaire des mécanismes impliqués habituellement dans la recombinaison génétique et/ou dans la synthèse d'ADN (tels que le crossing-over inégal ou la conversion génique). Les mécanismes responsables des translocations entre chromosomes non homologues demeurent inconnus.

Une puissante force de changement au sein des génomes eucaryotes et procaryotes est constituée par la capacité qu'ont certaines séquences de se déplacer d'un site à un autre. Ces séquences sont appelées **éléments transposables** ou **transposons**. La *transposition* ne repose pas sur une quelconque relation entre les

séquences du site donneur et du site receveur, différence importante comparée à tous les autres mécanismes impliqués dans la reconstruction de l'ADN.

Chaque transposon porte les gènes nécessaires à sa propre transposition bien qu'il puisse également utiliser des fonctions présentes dans le génome dans lequel il réside (telles que l'ADN polymérase et l'ADN gyrase). Chez les eucaryotes, la fréquence de tels événements peut être déduite d'après les similarités générales de structure et la diversité dans la localisation d'éléments analogues (voir Chapitre 30).

Quelles sont les conséquences de la présence de séquences transposables capables de se déplacer indépendamment vers de nouvelles localisations ? Un avantage possible pour l'organisme repose sur la propriété de la transposition de permettre des réarrangements directement ou indirectement. L'événement de transposition lui-même peut causer des délétions ou des inversions ou peut conduire au transport d'une séquence de l'hôte vers une nouvelle localisation.

Les transposons peuvent amener les systèmes cellulaires à fonctionner comme « des régions mobiles d'homologie » ; deux copies d'un transposon dans des positions différentes (même sur des chromosomes différents) peuvent fournir des sites pour la recombinaison réciproque. Il peut résulter de tels échanges des délétions, des insertions, des inversions ou des translocations.

Des éléments transposables ont effectivement joué un rôle actif dans l'évolution de certains plasmides bactériens. Les composants des plasmides ont pu être assemblés grâce à des réactions nécessitant la reconnaissance de séquences de transposon. L'intégration du plasmide F (sexuel) de *E. coli* dans le chromosome hôte s'effectue souvent par recombinaison réciproque dépendante de l'hôte (RecA), entre un transposon de l'ADN plasmidique et un transposon homologue situé dans l'ADN bactérien.

Pour ce qui est du transposon lui-même, les avantages conférés par ses activités intermittentes constituent une cible bien hypothétique pour la sélection naturelle. Ce point de vue suggère que les éléments transposables (du moins certains) peuvent ne conférer ni avantage, ni désavantage pour le phénotype, mais pourraient constituer de l'ADN égoïste uniquement concerné par sa propre propagation.

D'après ce concept, la relation du transposon au génome s'apparente à celle du parasite avec son hôte. La propagation des éléments par transposition est vraisemblablement contrebalancée par le tort entraîné si l'événement de transposition inactive un gène indispensable, ou si le nombre de transposons devient trop important pour les systèmes cellulaires. Nous devons,

par conséquent, nous souvenir que tout événement de transposition conférant un avantage sélectif — par exemple, un réarrangement génétique — va conduire prioritairement à la survie du génome portant le transposon actif.

LES SÉQUENCES D'INSERTION SONT DES TRANSPOSONS SIMPLES

Des éléments transposables ont été découverts sous forme d'insertions spontanées dans les opérons bactériens. Une telle insertion empêche la transcription et/ou la traduction du gène dans lequel elle est insérée. En comparant les séquences présentes chez différents mutants d'insertion, il est possible de classer plusieurs éléments distincts.

Les transposons les plus simples sont appelés **séquences d'insertion** (illustrant la façon dont elles ont été détectées). Chaque type est désigné par le préfixe IS, suivi d'un nombre qui l'identifie (les premières classes ont été numérotées IS1-4 ; les dernières classes ont des numéros qui reflètent l'histoire de leur isolement mais qui ne correspondent pas au nombre total d'éléments isolés à ce jour).

Les éléments IS sont des constituants normaux des chromosomes bactériens et des plasmides. Par exemple, une souche standard de *E. coli* possède huit copies de IS1 et cinq copies de IS2. Pour décrire une insertion dans un site particulier, un double « deux points » est utilisé ; ainsi, $\lambda :: IS1$ décrit un élément IS1 inséré dans le phage lambda.

Les éléments IS sont des unités autonomes, chacune d'entre elles code seulement pour les protéines nécessaires à sa propre transposition. Bien que tous les éléments IS diffèrent dans leurs séquences, ils partagent une même forme commune d'organisation. Les caractéristiques de quelques éléments IS courants sont résumées dans le **Tableau 29.1**.

Quand un élément IS se transpose, une séquence de l'ADN hôte est dupliquée au site d'insertion. La nature de la duplication est déterminée grâce à la comparaison des séquences du site accepteur avant et après l'insertion.

La **Figure 29.1** montre qu'au site de l'insertion, l'ADN IS est toujours bordé par de très courtes **répétitions directes** (dans ce contexte, « direct » signifie que deux copies d'une séquence sont répétées dans la même orientation et non pas que les répétitions sont adjacentes). Mais dans le gène d'origine (sans insertion), on trouve au site cible la séquence de l'une de ces répétitions. Dans la figure, le site accepteur est constitué par la séquence ATGCA TACGT. Après la transposition, une copie

Tableau 29.1
Certains transposons sont constitués ou contiennent des modules individuels.

Élément	Longueur	Répétitions terminales inversées	Répétitions directes au site cible	Protéines nécessaires à la transposition	Sélection du site cible
IS1	768 pb	23 pb	9pb	2	régionale
IS2	1 327 pb	41 pb	5pb	1	points chauds
IS4	1 428 pb	18 pb	11 ou 12 pb	1	AAAN ₂₀ TTT
IS5	1 195 pb	16 pb	4pb	1	points chauds
IS10R	1 329 pb	22 pb	9pb	1	points chauds
IS50R	1 531 pb	9 pb	9pb	1	points chauds
IS903	1 057 pb	18 pb	9pb	1	inconnue

Des transposons dont l'insertion se fait au hasard au sein d'une petite région peuvent montrer une préférence pour une région générale comparée à une autre. Par exemple, les insertions peuvent être plus fréquentes sur certains éléments particuliers de 3 kb bien que les sites soient toujours choisis au hasard au sein de ces éléments. On ne connaît pas les raisons de cette préférence qui est décrite comme un «profil» régional. La nature des points chauds n'est généralement pas connue. IS4 est un élément exceptionnel qui reconnaît une séquence cible particulière (dont l'un des brins est décrit).

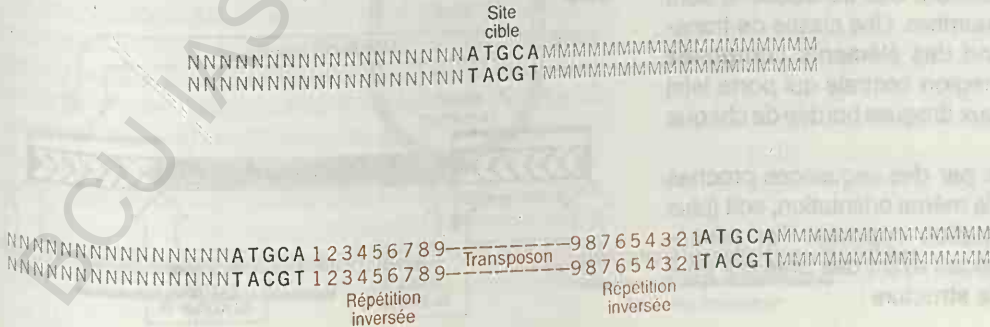
de cette séquence est présente à chaque extrémité du transposon.

La plupart des transposons s'insèrent dans des sites au hasard, au sein de l'ADN hôte. Certains montrent (à des degrés divers) des préférences pour des points chauds («hot spot»). Pour la plupart des transposons, la *séquence* de la répétition directe est différente pour chaque événement de transposition, mais la *longueur* est constante (à l'image du mécanisme de transposition). Les longueurs les plus fréquentes pour les répétitions directes sont de 5 et 9 pb.

Chaque élément se termine par de courtes **répétitions terminales inversées**. Habituellement, les répétitions sont de 15 à 25 pb et les deux copies sont étroitement apparentées plutôt qu'identiques. Comme le montre la Figure 29.1, la présence des répétitions terminales inversées signifie que la même séquence est rencontrée lorsque l'on progresse vers l'élément à partir de l'ADN bordant de l'une ou l'autre de ses extrémités.

Par conséquent, un transposon possède une structure caractéristique dans laquelle ses extrémités sont identifiées par les répétitions terminales inversées alors que les

Figure 29.1
Les transposons ont des répétitions terminales inversées et engendrent des répétitions directes dans l'ADN bordant au site cible.
Dans cet exemple, la cible a une séquence de 5 pb. Les extrémités du transposon sont constituées de répétitions inversées de 9 pb numérotées sur la séquence de 1 à 9.



extrémités adjacentes de l'ADN bordant sont caractérisées par de courtes répétitions directes. Détectée dans une séquence d'ADN, ce type d'organisation signe la présence d'un transposon et suggère de prime abord que la séquence provient d'un événement de transposition.

Un transposon est constitué d'une séquence linéaire constante. Par conséquent, toutes les copies d'un transposon donné ont les mêmes répétitions inversées terminales à la jonction avec l'ADN de l'hôte. *La reconnaissance des extrémités est commune à tous les événements de transposition.* Des mutations agissant en *cis*, empêchant la transposition, sont localisées aux extrémités qui sont reconnues par la (les) protéine(s) responsable(s) de la transposition. La protéine est appelée une **transposase**.

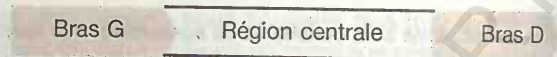
Le transposon le plus court, IS1, contient deux cadres de lecture ouverts dans la même direction ; ils sont traduits en protéines InsA et InsB, toutes deux nécessaires à la transposition. Tous les autres éléments IS contiennent une longue région codante unique, qui débute juste à l'intérieur de la répétition inversée de l'une des extrémités et qui se termine juste avant ou dans la répétition inversée de l'autre extrémité. Elle code pour la transposase.

La fréquence de transposition varie suivant les éléments. Les insertions dans des sites cibles distincts surviennent à une fréquence comparable au taux de mutation spontanée, habituellement entre 10^{-5} et 10^{-7} par génération. Puisque les insertions peuvent toucher plusieurs loci, la vitesse générale de transposition peut être d'environ 10^{-3} - 10^{-4} par élément et par génération. La réversion (par excision précise) est rare, les taux allant de 10^{-6} à 10^{-10} par génération, soit plus de mille à dix mille fois moins que l'insertion.

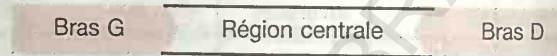
LES TRANSPOSONS COMPOSÉS ONT DES MODULES IS

Certains transposons portent des marqueurs de résistance aux drogues (ou autres) en plus des fonctions impliquées dans la transposition. Ces transposons sont dénommés **Tn**, suivi d'un nombre. Une classe de transposons plus gros comprend des **éléments composés** qui sont constitués d'une région centrale qui porte le(s) marqueur(s) de résistance aux drogues bordée de chaque côté par des **bras**.

Les bras sont constitués par des séquences proches qui peuvent être soit dans la même orientation, soit (plus fréquemment) dans l'orientation opposée. Par conséquent, un transposon composé ayant des bras qui sont des répétitions directes a la structure :



Si les bras sont des répétitions inversées, la structure est :



Les flèches indiquent le sens des bras qui sont représentés par G et D suivant une orientation (arbitraire) de la carte génétique du transposon de la gauche vers la droite. La structure d'un transposon composé est représentée de façon détaillée dans la **Figure 29.2**.

Les bras sont constitués par des éléments IS appelés quelquefois modules IS, par référence à leur participation au transposon composé. Puisque chaque module a la structure habituelle qui se termine par des répétitions inversées, le transposon composé se termine également par les mêmes courtes répétitions inversées. Les propriétés de quelques transposons composés sont résumées dans le **Tableau 29.2**.

Dans certains cas, les modules d'un transposon composé sont identiques comme dans Tn9 (répétitions directes de IS1), ou Tn903 (répétitions inversées de IS903). Dans d'autres cas, les modules sont apparentés mais ils ne sont pas identiques. Nous pouvons alors distinguer les modules tels que IS10L et IS10R.

Un module IS fonctionnel peut être capable de se transposer lui-même, comme le montrent les exemples de IS10R de Tn10 ou de IS50R de Tn5. Lorsque les modules d'un transposon composé sont identiques, l'un ou l'autre peut probablement permettre le mouvement du transposon. Lorsque les modules sont différents, ils

Figure 29.2

Un transposon composé possède une région centrale qui porte des gènes sans rapport avec la transposition (comme la résistance aux drogues), et qui est flanquée par des modules IS. Les modules ont de courtes répétitions terminales inversées. Si les modules eux-mêmes sont inversés (comme il est dessiné), les courtes répétitions terminales aux extrémités du transposon sont identiques.

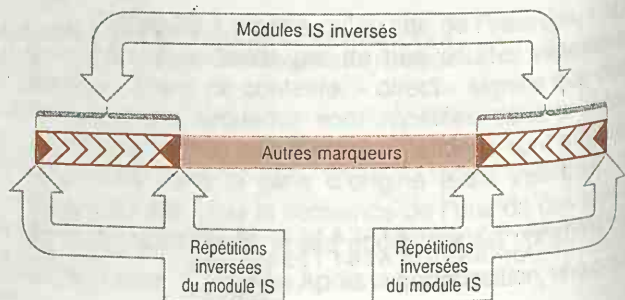


Tableau 29.2
Certains transposons sont composés et sont entourés par des modules individuels.

Élément	Longueur (pb)	Marqueurs génétiques	Modules terminaux	Orientation des modules	Relation entre modules	Fonctions du module
Tn10	9 300 pb	<i>tet^R</i>	IS10R IS10L	inversée	2,5 % divergence	complètement fonctionnel fonction réduite
Tn5	5 700 pb	<i>kan^R</i>	IS50R IS50L	inversée	changement 1 pb	fonctionnel non fonctionnel
Tn903	3 100 pb	<i>kan^R</i>	IS903	inversée	identiques	tous deux fonctionnels
Tn9	2 500 pb	<i>cam^R</i>	IS1	directe	supposés identiques	supposés fonctionnels

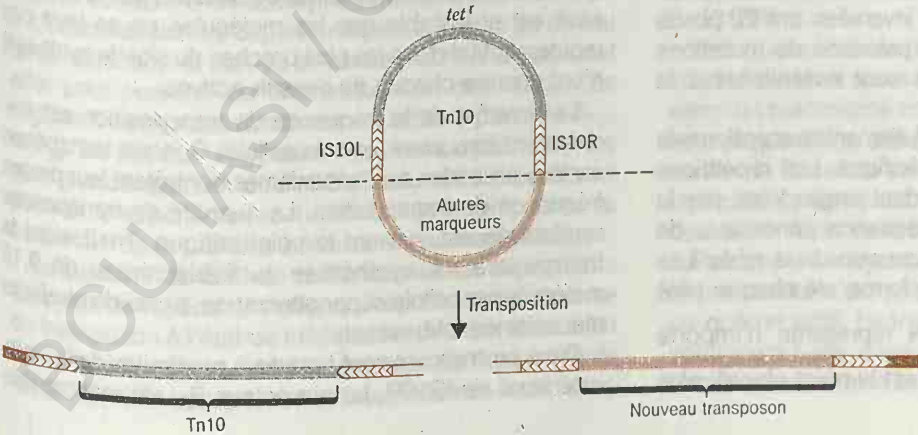
peuvent varier dans leur capacité fonctionnelle de telle sorte que la transposition peut dépendre complètement ou principalement de l'un des modules.

On pense que les transposons composés proviennent de l'association de deux modules, au départ indépendants, avec la région centrale. Un tel événement peut se produire quand un élément IS se transpose à un site receveur proche du site donneur. Tout d'abord identiques, les modules peuvent demeurer inchangés mais peuvent aussi diverger. La capacité d'un seul module à transposer un élément composé entier peut expliquer la perte de pression sélective pour le maintien de l'activité des deux modules.

Nous aimerions savoir pourquoi s'effectue la transposition du transposon composé à la place de la seule transposition du module individuel. Cette question est particulièrement importante dans les cas où les deux modules sont fonctionnels. Par exemple Tn9, où les modules sont des éléments IS1 et sont certainement aussi actifs l'un que l'autre sur le transposon composé. Pourquoi le transposon est-il préservé dans sa totalité alors que chaque séquence insérée pourrait avoir une existence propre ?

Deux éléments IS sont en fait capables de transposer toute séquence se trouvant entre eux, tout comme ils sont capables de se transposer eux-mêmes. La Figure

Figure 29.3
Les modules IS10 peuvent coopérer afin de transposer toute région d'ADN localisée entre eux. Lorsqu'une petite molécule circulaire porte Tn10, les répétitions IS10 peuvent transposer l'une ou l'autre partie du cercle. La transposition de *tet^R* correspond au mouvement de Tn10. La transposition des marqueurs situés à l'opposé crée un nouveau transposon « inside-out ».



29.3 montre que si Tn10 se trouve sur un réplicon circulaire, on considère que ses deux modules bordent soit le gène *tet^R* du Tn10 d'origine, soit la séquence de l'autre partie du cercle. Un événement de transposition peut donc mettre en jeu soit le transposon Tn10 d'origine, soit le transposon formé avec la région centrale opposée.

Il faut remarquer que les deux transposons ont des modules inversés mais que ceux-ci peuvent évidemment fonctionner dans n'importe quelle direction par rapport à la région centrale. La fréquence de transposition des transposons composés diminue avec la distance existant entre les modules. La dépendance de la longueur a pu être un facteur déterminant dans les tailles des transposons composés courants. La force majeure gouvernant la transposition des transposons composés a pu être la sélection pour le(s) marqueur(s) porté(s) par la région centrale.

Les deux produits de IS1 et le produit unique de tous les autres éléments sont probablement responsables de la création d'un site cible et de la reconnaissance des extrémités du transposon. *Pour que le transposon serve de substrat à la transposition, seules les extrémités sont nécessaires.* Nous savons peu de choses sur les détails du processus et sur la chronologie des événements impliqués dans ces réactions, mais nous supposons qu'ils sont du même type pour tous les éléments IS et les transposons composés.

UN SEUL MODULE DE Tn10 EST FONCTIONNEL

Les rapports existants entre les modules qui ne sont plus identiques peuvent donner un aperçu de l'évolution du transposon composé. Les exemples de Tn10 et Tn5 sont les mieux caractérisés.

Seulement quelques bases sont nécessaires à la reconnaissance de l'élément comme substrat de la transposition. Les répétitions terminales inversées ont 22 pb de longueur. Comme le montre la présence de mutations agissant *in cis*, les extrémités sont essentielles à la réaction de transposition.

Tn10 est l'un de ces (quelques) éléments exceptionnels qui ont une séquence cible spécifique. Les répétitions directes de 9 pb de l'ADN bordant engendrées par la transposition possèdent une séquence consensus de 6 pb disposée symétriquement au sein de la cible. Les répétitions ont donc souvent la forme, de chaque côté de Tn10, NGCTNAGCN, NCGANTCGN, où N représente n'importe quelle paire de bases. Plus fort est le point chaud, plus

sa conformation est proche de la séquence consensus. Il est probable que ce soit la fonction même codée par le transposon qui reconnaisse la séquence cible et les extrémités du transposon.

L'élément IS10R est le module actif du transposon Tn10. Le module IS10L est fonctionnellement défectueux et n'a que 1 à 10 % de l'activité de transposition de IS10R. L'accumulation des mutations dans IS10L par rapport à IS10R correspond à environ 2,5 % de divergence entre les modules.

L'organisation de IS10R est résumée dans la Figure 29.4. Il y a sur un brin un cadre de lecture continu ; la signification de l'autre cadre de lecture chevauchant n'est pas connue. Le niveau de la transposase est augmenté par des mutations « à haute fréquence de sauts » (« high-hopper ») qui accroissent l'expression du gène de IS10R. La quantité de cette protéine peut donc limiter la vitesse de transposition.

Les mutants de ce gène sont difficilement complémentés *in trans* par un autre élément de type sauvage. Ceci montre que la transposase agit préférentiellement *in cis* ; l'enzyme ne fonctionne efficacement qu'avec la matrice d'ADN à partir de laquelle elle a été transcrite et traduite. La préférence pour l'action *in cis* est une caractéristique commune aux protéines impliquées dans la transposition dépendante de IS (nous avons déjà décrit dans le Chapitre 14 une telle préférence pour une autre fonction, la protéine A intervenant dans la réplication de ϕ X174).

Lorsque la distance augmente entre le gène de la transposase de IS10 et les extrémités du transposon, l'efficacité de la transposase diminue. La raison de cette action préférentielle *in cis* est donc que la transposase reconnaît plus efficacement une séquence d'ADN cible lorsque celle-ci réside au plus près du lieu de synthèse de l'enzyme. Soit la transposase se lie à l'ADN tellement vite après (ou même pendant) la synthèse protéique qu'elle a une faible probabilité de diffuser, soit la transposase est si instable que les molécules ne se lient pas rapidement (et donc les plus proches du site de synthèse) n'ont aucune chance de devenir actives.

Le contrôle de la fréquence de transposition est évidemment important pour la cellule. Chaque transposon semble posséder des mécanismes contrôlant leur propre fréquence de transposition. La quantité de transposase synthétisée est souvent le point critique. Tn10, dont la transposase est synthétisée au faible niveau de 0,15 molécule par cellule et par génération, possède plusieurs mécanismes intéressants.

Deux promoteurs sont trouvés à proximité de la limite extérieure de IS10R. Le promoteur P_{IN} est responsable

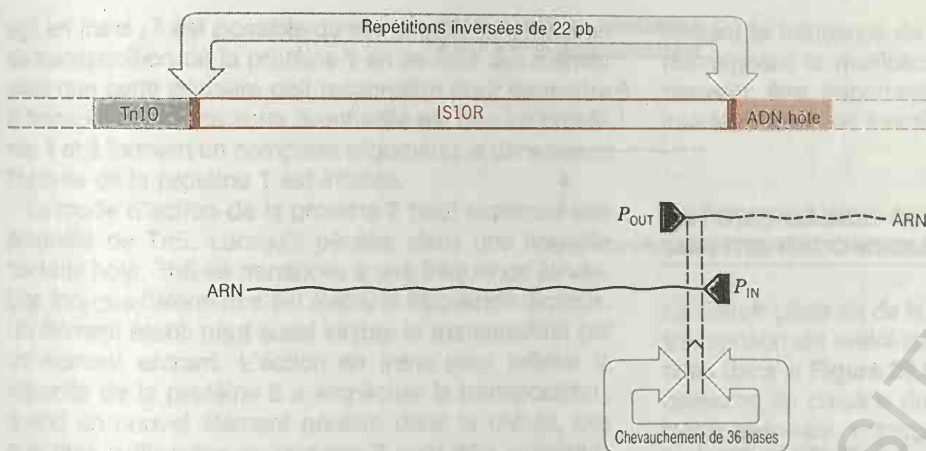


Figure 29.4

Deux promoteurs orientés à l'opposé l'un de l'autre se trouvent près de la frontière avec l'ADN hôte. Le promoteur fort P_{OUT} gouverne la transcription de l'ADN hôte bordant. Un ARN codant qui s'étend sur la longueur de IS10R est transcrit à partir du promoteur plus faible P_{IN} ; il correspond à une protéine de 47 000 daltons. Les transcrits « OUT » et « IN » se chevauchent sur 40 bases.

de la transcription de IS10R. Le promoteur P_{OUT} entraîne la poursuite de la transcription de l'ADN jusque dans les séquences bordantes adjacentes, effet qui serait responsable de l'activation des gènes bactériens adjacents à Tn10.

Le phénomène de l'inhibition par les copies multiples montre que l'expression du gène IS10R est régulée. La transposition sur le chromosome bactérien d'un élément Tn10 est réduite lorsque des copies supplémentaires de IS10R sont introduites par un plasmide multicopies. L'inhibition exige la présence du promoteur P_{OUT} ; et elle n'est effective que sur un gène IS10R exprimé à partir du promoteur P_{IN} (l'effet est perdu si le gène est placé sous le contrôle d'un autre promoteur).

L'inhibition par les copies multiples s'exerce au niveau de la traduction et non de la transcription. A la base, cet effet peut provenir du chevauchement des 40 pb des régions 5' terminales des transcrits initiés à partir de P_{OUT} et P_{IN} . Du fait que P_{OUT} est un promoteur plus fort, une plus grande quantité d'ARN OUT peut être présente ; il pourrait s'apparier avec l'ARN IN pour empêcher la traduction de la protéine nécessaire à la transposition.

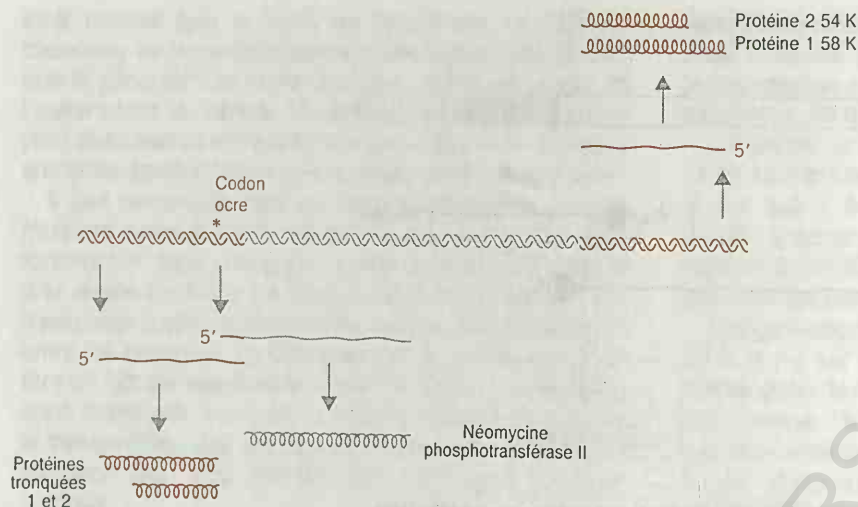
La capacité de IS10 de se transposer est dépendante du cycle de réplication, effet qui repose sur la réponse du transposon à l'état de méthylation de deux sites. L'un est à l'intérieur de la répétition inversée de l'extrémité de IS10R, là où la transposase est supposée se lier. L'autre

site est dans le promoteur P_{IN} à partir duquel est transcrit le gène de la transposase.

Ces deux sites sont méthylés par le système *dam* décrit dans le Chapitre 15. La méthylase Dam modifie l'adénine dans la séquence GATC sur un brin nouvellement synthétisé, engendré par la réplication. La fréquence de transposition de Tn10 est augmentée de 1 000 fois dans les souches *dam*⁻ chez lesquelles les deux sites cibles sont dépourvus de groupes méthyle. On suppose que dans la cellule sauvage, le passage de la fourche de réplication par ces sites engendre des séquences non méthylées et active le transposon en combinant une transcription plus fréquente du gène de la transposase à partir de P_{IN} , et une activation de la liaison de la transposase à l'extrémité de IS10R.

Pourquoi serait-ce si souhaitable que la transposition intervienne si tôt après la réplication ? Tn10 se transpose selon un mécanisme conservatif dans lequel l'ADN donneur risque d'être détruit (voir plus loin). Les chances de survie de la cellule peuvent être très augmentées si la réplication vient juste d'avoir lieu afin de fournir une deuxième copie de la séquence donneuse.

Puisque le transposon choisit son site cible au hasard, la probabilité n'est pas négligeable qu'il s'intègre dans un opéron actif. La transcription issue de l'extérieur se continuera-t-elle dans le transposon, activant ainsi la transposase, dont la surproduction peut en retour induire

**Figure 29.5**

Le changement d'une seule paire de bases dans IS50L entraîne la terminaison prématurée de la synthèse des protéines 1 et 2 et la création d'un promoteur pour la transcription du gène responsable de la résistance à la néomycine/kanamycine.

des niveaux élevés (voir l'étaux) de transposition ? Tn10 se protège de tels événements par deux mécanismes. La transcription à travers l'extrémité IS10R diminue son activité en inhibant probablement sa capacité de lier la transposase. De plus, l'ARNm qui prend son origine en amont du promoteur est faiblement traduit car il possède une structure secondaire dans laquelle le codon d'initiation est inaccessible.

LES MODULES DE Tn5 SONT PRESQUE IDENTIQUES MAIS TRÈS DIFFÉRENTS

Les modules de Tn5 sont un exemple frappant de la capacité qu'ont des mutations limitées de produire des effets fonctionnels majeurs. Le module IS50R est fonctionnel ; le module IS50L est non fonctionnel pour la transposition. L'unique différence entre ces deux modules réside dans la substitution d'une seule paire de bases. Les effets de cette substitution sont décrits dans la Figure 29.5.

Deux protéines sont produites à partir du même cadre de lecture dans IS50R. La seule différence est que la protéine 1 a environ 40 acides aminés supplémentaires du côté N-terminal qui sont absents de la protéine 2. Les événements précis impliqués dans la transcription et la

traduction de ces deux protéines sont encore à déterminer ; tout ce que l'on sait de façon certaine est que la protéine 2 est plus produite que la protéine 1.

Le changement d'une seule paire de bases dans IS50L affecte simultanément la traduction de ces protéines et le contrôle de la transcription de la région centrale. La substitution crée un codon ochre qui entraîne la terminaison prématurée de la traduction des protéines 1 et 2. Les protéines tronquées (parfois appelées 3 et 4) n'ont pas d'activité de transposition. La même substitution entraîne également la création d'un promoteur pour la transcription du gène de la région centrale qui code pour la néomycine phosphotransférase II, enzyme responsable de la résistance aux antibiotiques tels que la néomycine et la kanamycine. Le changement est donc nécessaire à cette fonction du transposon.

Les fonctions des protéines 1 et 2 sont voisines mais différentes. La protéine 1 est essentielle à la transposition, soit du module IS50, soit du transposon Tn5 intact. La protéine de type sauvage complémente très faiblement en *trans* un module défectif. La protéine 1 peut donc être le type même de fonction agissant en *cis* sur la transposition.

La protéine 2 inhibe la transposition. Son action s'exerce lors de quelque étape inconnue du processus de transposition lui-même (c'est-à-dire n'intervenant pas sur la régulation de l'expression génique). La protéine 2

agit en *trans* ; il est possible qu'elle empêche la fonction de transposition de la protéine 1 en se liant aux mêmes sites que cette dernière doit reconnaître pour permettre la transposition. Une autre éventualité est que les protéines 1 et 2 forment un complexe oligomérique dans lequel l'activité de la protéine 1 est inhibée.

Le mode d'action de la protéine 2 peut expliquer une propriété de Tn5. Lorsqu'il pénètre dans une nouvelle bactérie hôte, Tn5 se transpose à une fréquence élevée. Une fois que l'élément s'est établi, la fréquence diminue. Un élément établi peut aussi inhiber la transposition par un élément entrant. L'action en *trans* peut refléter la capacité de la protéine 2 à empêcher la transposition ; quand un nouvel élément pénètre dans la cellule, des quantités suffisantes de protéine 2 sont déjà présentes pour inhiber l'action en *cis* de la protéine 1 du nouvel élément mais aussi de l'ancien élément Tn5. Les protéines tronquées de IS50L pourraient être également impliquées dans cet effet.

Le rapport existant entre la protéine 2 et la protéine 1 de Tn5 est naturellement différent de celui existant entre les transcrits de P_{OUT} et P_{IN} de Tn10. Ces interactions partagent toutefois une caractéristique commune : elles

limitent la fréquence de transposition. Des mécanismes restreignant la multiplication à l'infini des transposons peuvent être importants pour les empêcher de trop interférer avec les fonctions cellulaires.

RECOMBINAISON CONSERVATIVE CONTRE RECOMBINAISON RÉPLICATIVE

La nature générale de la réaction au cours de laquelle un transposon est inséré dans un nouveau site est schématisée dans la **Figure 29.6**. Elle consiste en la création de cassures en chicane dans l'ADN cible, en la liaison du transposon aux extrémités monobrin sortantes et au remplissage des brèches. La création et le remplissage des extrémités en chicane explique l'apparition des répétitions directes de l'ADN cible au site d'insertion. Le décalage entre les coupures sur les deux brins détermine la longueur des répétitions directes ; par conséquent, la répétition cible caractéristique de chaque transposon reflète la géométrie de l'enzyme impliquée dans la coupeure de l'ADN cible.

On peut envisager deux types généraux de processus pour la transposition, selon le devenir du site donneur :

- L'élément se transposant peut être conservé, il se déplace d'un site à un autre comme une entité physique.

La **Figure 29.7** décrit les conséquences d'une transposition conservative. L'élément va directement du site donneur au site receveur. L'élément s'insère au site cible. Mais que pourrait-il se produire au site donneur ? Le dommage causé par la perte de l'élément devrait être en principe réparé. Ou le répliqueur donneur pourrait être détruit par le fait même de la transposition.

Figure 29.6
Les répétitions directes de l'ADN cible bordant le transposon sont engendrées par la création de coupures en chicane dont les extrémités sortantes sont liées au transposon.

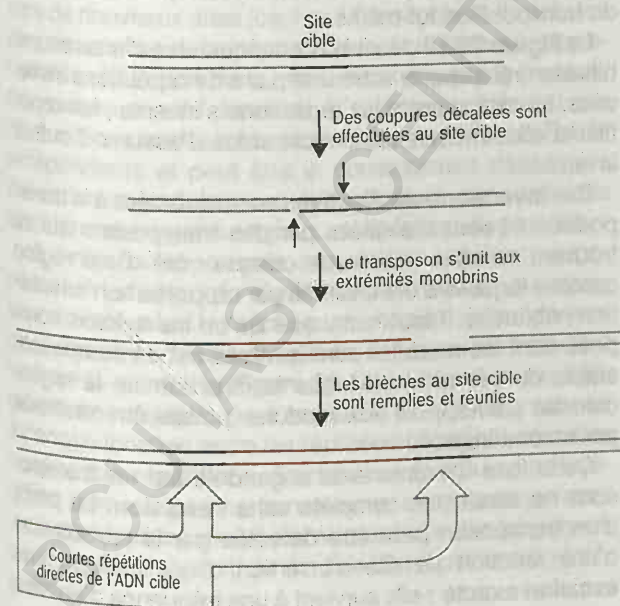


Figure 29.7

La transposition conservative permettrait le déplacement du transposon en tant qu'entité physique d'un site donneur à un site receveur. Ce qui se produirait au site donneur n'est pas clair.



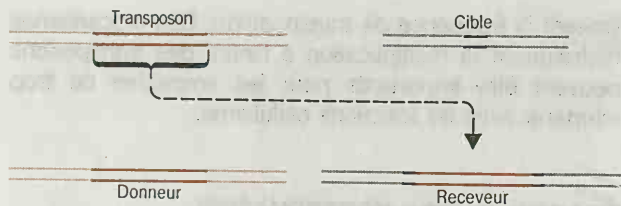


Figure 29.8

La transposition répliquative conduit à la création d'une copie du transposon qui s'insère au site donneur ; le site receveur demeure inchangé.

- L'élément peut être dupliqué pendant la réaction de telle sorte que l'entité se transposant soit une copie de l'élément d'origine.

La Figure 29.8 décrit les événements résultant d'une transposition répliquative. Le transposon est copié pendant son déplacement. Une copie reste au site d'origine tandis que l'autre s'insère au nouveau site. La transposition s'accompagne donc d'une augmentation du nombre de copies du transposon.

En plus de la transposition intermoléculaire « simple » qui conduit à une insertion dans un site nouveau, des transposons peuvent entraîner d'autres types de réarrangements de l'ADN. Certains de ces événements sont les conséquences des rapports existant entre les copies multiples du transposon. D'autres représentent les issues possibles du mécanisme de transposition et donnent quelques indices sur la nature des événements sous-jacents. De tels événements fournissent la meilleure information sur l'existence de différents mécanismes de la transposition.

Des réarrangements dans l'ADN hôte peuvent résulter de l'insertion d'une copie du transposon à un site proche de celui d'origine. Les systèmes hôtes peuvent réaliser une recombinaison réciproque entre les deux copies du transposon ; les conséquences dépendent de l'orientation, identique ou inversée des répétitions.

La Figure 29.9 illustre la règle générale selon laquelle la recombinaison entre toute paire de répétitions directes entraîne la délétion du matériel situé entre celles-ci. Ce matériel est excisé sous forme d'un ADN circulaire (qui est perdu par la cellule) ; le chromosome garde une copie unique de la répétition directe. Il faut noter que la recombinaison entre les modules IS1 directement répétés, du transposon composé Tn9, conduirait par conséquent à un transposon avec un seul module IS1.

La délétion des séquences adjacentes à un transposon

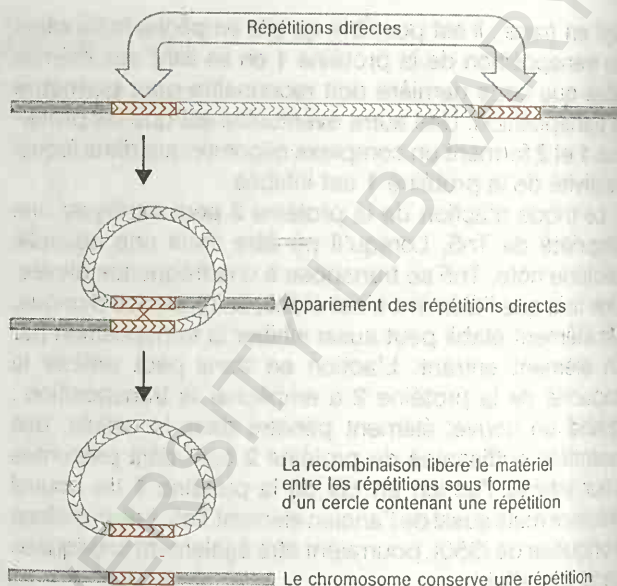


Figure 29.9

La recombinaison réciproque entre les répétitions directes conduit à l'excision du matériel compris entre ces dernières ; chaque produit de la recombinaison contient une copie de la répétition directe.

pourrait donc résulter d'un processus en deux étapes ; la transposition engendre une répétition directe d'un transposon et la recombinaison s'effectue entre les répétitions. Pourtant, la majorité des délétions qui ont lieu au voisinage des transposons proviennent probablement d'une variation dans le déroulement suivi par l'événement de transposition lui-même.

La Figure 29.10 décrit les conséquences d'une recombinaison réciproque entre une paire de répétitions inversées. La région entre les répétitions s'inverse ; les répétitions elles-mêmes restent capables d'assurer d'autres inversions.

Des **inversions duplicatives** sont associées à la transposition et sont identifiées par des transposons qui se trouvent en sens inverse de chaque côté d'une région centrale laquelle a été inversée par rapport à son orientation habituelle. Il faut remarquer qu'un transposon composé dont les modules sont inversés est un composant stable du génome, bien que la direction de la région centrale par rapport aux modules puisse être inversée par recombinaison.

Cette liste d'événements engendrés par les transposons ne serait pas complète sans l'**excision**. La perte d'un transposon peut être détectée par la restauration d'une fonction au site d'insertion. Ceci implique une **excision exacte** ; elle survient à une fréquence d'environ

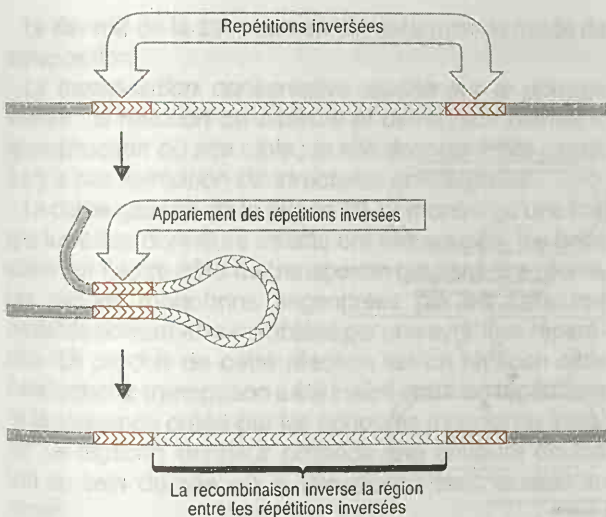


Figure 29.10
La recombinaison réciproque entre les répétitions inversées entraîne l'inversion de la région localisée entre celles-ci.

10^{-6} pour Tn5. La perte peut être aussi détectée par un retour d'une polarité des gènes adjacents ; ceci peut être la conséquence d'une **excision inexacte** (qui laisse un vestige du transposon, suffisant pour empêcher la réactivation du gène cible mais inapte à causer des effets polaires). L'excision inexacte survient à une fréquence d'environ 10^{-5} pour Tn5.

Aucun type d'excision n'est associé à la transposition vers de nouveaux sites (ceci va à l'encontre de l'existence de mécanismes de transposition conservative qui assureraient la réparation de la cassure causée par la perte du transposon au site donneur). L'excision ne dépend pas de fonctions codées par le transposon ; elle est RecA-indépendante et peut être la conséquence de certains mécanismes cellulaires qui engendrent des délétions spontanées entre des séquences répétées très proches.

Une réaction importante, assurée par certains transposons, est la **fusion de réplicons** pour former une structure **co-intégrée**. Un réplicon contenant un transposon peut fusionner avec un réplicon dépourvu de l'élément comme le montre la **Figure 29.11**. La structure co-intégrée qui en résulte a deux copies du transposon, soit une à chaque jonction entre les réplicons d'origine, orientées comme des répétitions directes. Les structures co-intégrées formées par les transposons composés peuvent avoir une duplication soit de l'élément en entier (structures co-intégrées Tn), soit du module IS (structures co-intégrées IS).

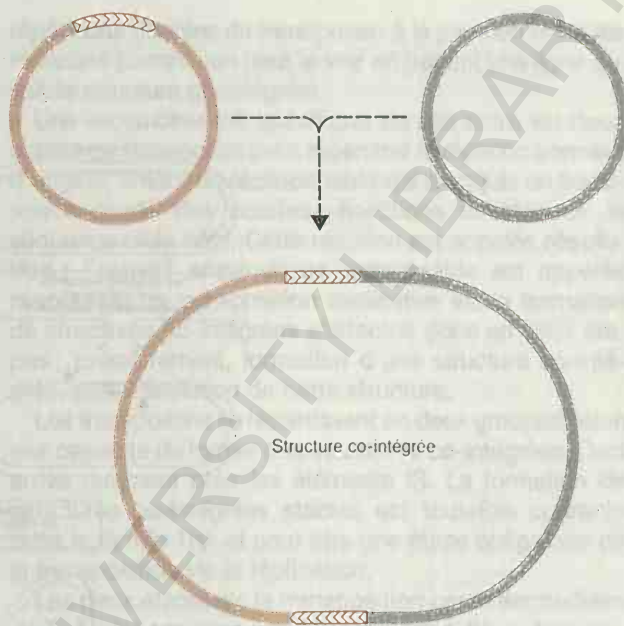


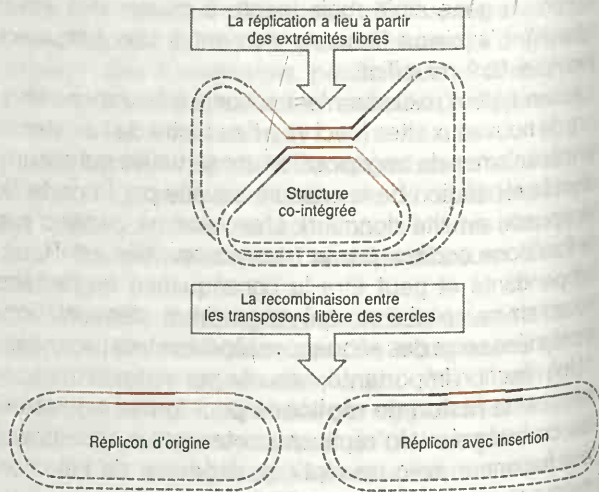
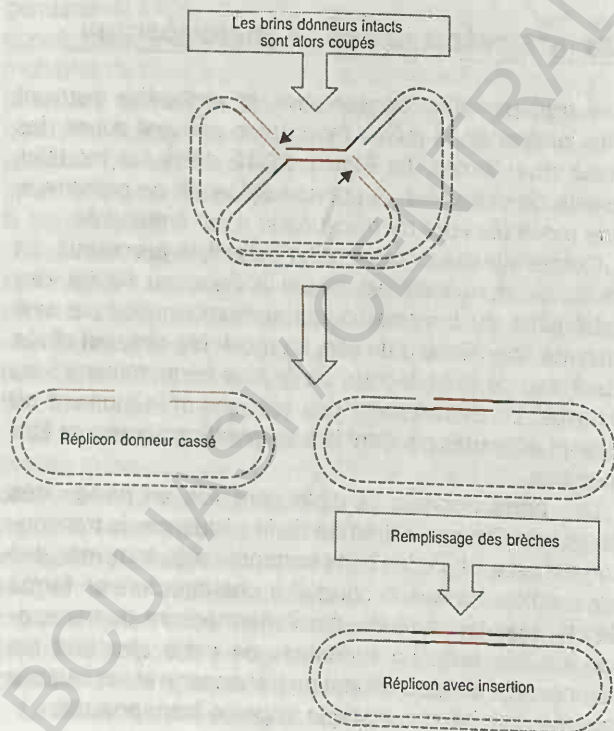
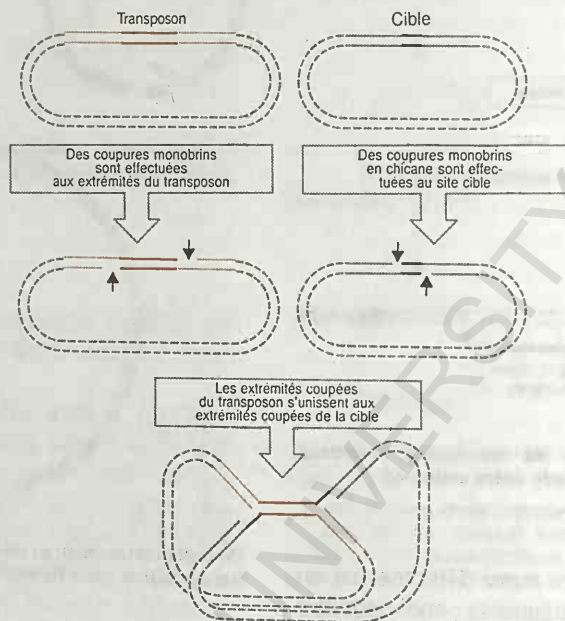
Figure 29.11
Un réplicon donneur et un réplicon receveur peuvent fusionner par transposition pour former une structure co-intégrée.

LES INTERMÉDIAIRES DE LA TRANSPPOSITION

Les transpositions conservative et répllicative peuvent être initiées de la même façon puis peuvent suivre des voies divergentes. La **Figure 29.12** décrit les modèles usuels de ces voies. *A aucune étape de ce processus, une molécule libre de transposon n'est engendrée.*

Quatre coupures monobrin initient le processus. La molécule donneuse est clivée à l'une ou l'autre des extrémités du transposon qui sont reconnues par une enzyme spécifique d'un site. La molécule cible est clivée au niveau de sites décalés de (le plus fréquemment) 5 ou 9 bases. Le déroulement des clivages et le moment où ils sont effectués pendant la transposition pourraient être modifiés.

Les brins donneur et cible sont liés au niveau des coupures. Chaque extrémité de la séquence du transposon est unie à l'un des brins sortants créés au niveau des sites cibles. La liaison conduit à une structure en forme de chi, maintenue réunie par l'intermédiaire du transposon double brin. La formation de cette structure est responsable du mouvement du transposon et est assurée par une activité enzymatique appelée **transposase**.



Le devenir de la structure en *chi* détermine le mode de transposition.

La *transposition conservative* repose sur le principe suivant : la réaction de cassure et de réunion permet la reconstruction du site cible ; le site donneur reste cassé. Il n'y a pas formation de structures co-intégrées.

La partie gauche de la Figure 19.12 montre qu'une fois que les brins donneurs intacts ont été coupés, les brins cibles sur l'autre côté du transposon peuvent être réunis. Les régions monobrins engendrées par les coupures décalées doivent être comblées par une synthèse réparatrice. Le produit de cette réaction est un réplicon cible dans lequel le transposon a été inséré entre les répétitions de la séquence créée par les coupures monobrins initiales. Le réplicon donneur possède une coupure double brin au sein du site où le transposon était localisé au départ.

Ce peut être le mécanisme utilisé pour la transposition de l'élément IS ; une transposase unique (ou la combinaison de deux produits dans le cas de IS1) est probablement responsable des deux étapes de la réaction. Nous ne connaissons pas encore les détails des réactions moléculaires assurées par la transposase. La formation occasionnelle de structures co-intégrées par des éléments IS pourrait avoir lieu dans les cas rares où la réaction s'engage vers l'autre voie.

Le principe de la transposition répllicative est que la réplication entraîne la duplication du transposon et, par conséquent, la création de copies aux sites donneur et cible. Le produit est une structure co-intégrée.

La partie droite de la Figure 29.12 montre comment le transposon pourrait être dupliqué. La structure en *chi* contient une région simple brin à chacune des extrémités en chicane. Ces régions sont des pseudo-fourches de réplication qui fournissent une matrice pour la synthèse d'ADN (l'utilisation des extrémités en tant qu'amorces de réplication implique que la coupure des chaînes doit s'effectuer avec une polarité qui engendre une extrémité 3'-OH à ce niveau).

Si la réplication se fait à partir de l'une ou l'autre de ces pseudo-fourches de réplication, elle se poursuivra le long du transposon en séparant ses brins et se terminera à ses extrémités. La réplication est probablement assurée par des fonctions codées par l'hôte. A ce stade, la structure est une structure co-intégrée qui possède des

répétitions directes du transposon à la jonction entre les réplicons (comme on peut le voir en traçant une ligne qui suit la structure co-intégrée).

Une recombinaison spécifique de site entre les deux copies du transposon peut régénérer le réplicon donneur d'origine, créant un réplicon cible qui a acquis un transposon bordé des courtes répétitions directes de la séquence cible hôte. Cette réaction est appelée **résolution** ; l'activité enzymatique responsable est appelée **résolvase**. La transposition répllicative via la formation de structures co-intégrées s'effectue donc en deux étapes : premièrement, formation d'une structure co-intégrée, puis, résolution de cette structure.

Les transposons se répartissent en deux groupes selon leur capacité de former des structures co-intégrées. Ceci arrive rarement pour les éléments IS. La formation de structures co-intégrées stables est toutefois courante dans la famille TnA et peut être une étape obligatoire de la transposition via la réplication.

Les deux étapes de la transposition par l'intermédiaire de TnA sont assurées par différents produits du transposon. L'étape de transposition met en jeu les extrémités de l'élément, tout comme cela est fait chez les éléments de type IS. La résolution nécessite un site interne spécifique, caractéristique propre à la famille TnA.

La famille TnA est constituée par des grands transposons (environ 5 kb) qui ne sont pas composés par des modules de type IS mais qui comprennent des unités indépendantes ; celles-ci portent des gènes codant pour la transposition et pour d'autres propriétés telle que la résistance aux drogues. La famille TnA est constituée de nombreux transposons apparentés, certains d'entre eux sont décrits dans le **Tableau 29.3**.

Les transposons TnA ont la caractéristique terminale habituelle qui consiste en des répétitions inversées très proches généralement d'environ 38 pb de longueur. Des délétions agissant en *cis*, dans l'une ou l'autre des répétitions, empêchent la transposition d'un élément. Une répétition directe de 5 pb est engendrée au site cible, qui est choisi au hasard, quoique avec une préférence pour une région.

Les deux gènes dont les produits sont nécessaires à la transposition ont été identifiés par des mutations récessives. Puisque les gènes mutants peuvent être complétés en *trans* par des gènes sauvages appar-

Figure 29.12
Les transpositions conservative et répllicative pourraient partager un intermédiaire commun. Une insertion conservative peut résulter de la libération par coupure de la structure en *chi*. Si le transposon de la structure en *chi* est répliqué, il y a formation de structure co-intégrée ; sa résolution par recombinaison conduit à une transposition répllicative.

Tableau 29.3

Les membres de la famille des transposons TnA ont la même organisation générale mais ils peuvent porter différents marqueurs génétiques.

Élément	Marqueurs génétiques	Répétitions terminales inversées	Répétition bordante au site cible	Sélection du site cible
Tn3	<i>amp^R</i>	38 pb	5 pb	régionale
Tn1	<i>amp^R</i>	?	5 pb	régionale
Tn1000 ($\gamma\delta$)	?	37 pb	5 pb	régionale
Tn501	<i>Hg^R</i>	38 pb	5 pb	inconnue

tenant à une autre copie du transposon leurs actions ne montrent donc pas la préférence en *cis* caractéristique du processus de transposition de type IS.

Les mutants dans *tnpA* ne peuvent pas se transposer. Le produit du gène est une transposase qui se lie à l'ADN monobrin. On connaît peu de choses sur sa fonction ; elle est probablement nécessaire à la reconnaissance des extrémités de l'élément et elle effectuerait des coupures décalées de 5 pb dans l'ADN cible auquel se lie le transposon.

Les transposases de Tn3 et de l'élément apparenté Tn1000 (appelé autrefois $\gamma\delta$) ne fonctionnent avec efficacité que sur leurs propres transposons bien que les extrémités de ces deux éléments soient identiques sur 27 des 38 pb des répétitions inversées. Certaines de ces différences doivent être critiques pour la reconnaissance des extrémités.

Les fonctions doubles du produit du gène *tnpR* ont été mises en évidence par les divers effets de mutations. La protéine agit à la fois comme un répresseur de l'expression génique et comme une résolvase.

Les mutations dans *tnpR* augmentent la fréquence de transposition. La raison est que TnpR réprime la transcription de *tnpA* et de son propre gène. L'inactivation de la protéine TnpR permet donc une synthèse accrue de TnpA, ce qui a pour conséquence d'augmenter la fréquence de transposition. Ceci implique que la quantité de transposase TnpA doit être un facteur limitant dans la transposition.

Les gènes *tnpA* et *tnpR* sont exprimés de façon divergente à partir d'une région de contrôle intercistronique riche en A-T, comme l'indique la carte de Tn3 donnée dans la Figure 29.13. Tous les effets de TnpR se font par l'intermédiaire de sa liaison à cette région.

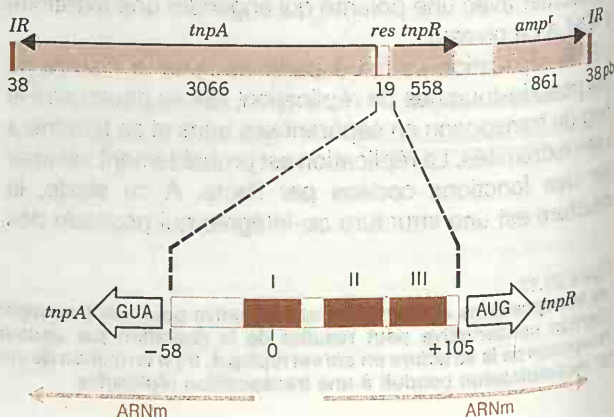
De par sa propriété de résolvase, TnpR est impliquée dans la recombinaison entre les répétitions directes de Tn3 dans la structure intégrée. La structure co-intégrée est un intermédiaire obligatoire de la transposition ; la résolution est bloquée chez les mutants *tnpR* qui ne peuvent donner lieu qu'à des structures co-intégrées.

Une structure co-intégrée peut être en principe résolue par recombinaison homologue entre toute paire corres-

Figure 29.13

Des transposons comme Tn3 ont des répétitions terminales inversées, un site *res* interne et trois gènes connus.

Les flèches indiquent la direction de l'expression des régions codantes. Les nombres du schéma supérieur indiquent la longueur en paires de bases de chaque région. Les nombres du schéma inférieur donnent la distance en paires de bases à partir du point de croisement (0).



pondante de points dans les deux copies du transposon (voir Figure 29.12). Mais la réaction de résolution de Tn3 ne peut s'effectuer qu'à un site spécifique.

Le site de la résolution est appelé *res*. Il a été identifié par des délétions agissant en *cis* qui empêchent l'achèvement de la transposition aboutissant à l'accumulation des structures co-intégrées. En l'absence de *res*, la réaction de résolution peut être effectuée par la recombinaison RecA-dépendante, mais ceci est beaucoup moins efficace.

Les sites où se lie la résolvasse TnpR ont été déterminés par empreinte du complexe ADN-protéine. Leurs localisations sont résumées dans la partie inférieure de la Figure 29.13. La liaison s'effectue de façon indépendante à chacun des trois sites tous longs de 30 à 40 pb. Les trois sites de liaison partagent tous une homologie de séquence qui définit la séquence consensus ayant une symétrie par rapport à un point.

Le site I comprend la région définie génétiquement comme étant le site *res* ; en son absence, la réaction de résolution ne s'effectue pas du tout. Toutefois, la résolution implique également la liaison aux sites II et III car la réaction s'effectue faiblement si l'un de ces sites est délété. Le site I semble chevaucher l'origine de la transcription de *tnpA*. Le site II chevauche l'origine de la transcription de *tnpR* ; une mutation dans l'opérateur est localisée juste à l'extrémité gauche du site.

Est-ce que ces sites interagissent ? Il est possible que la liaison aux trois sites soit nécessaire au maintien de l'ADN dans une topologie appropriée ; la réaction de résolution peut être déclenchée quand le complexe ADN-protéine d'un transposon interagit avec celui d'un autre transposon. La liaison à un groupe unique de sites peut réprimer la transcription de *tnpA* et de *tnpR* sans introduire de modification dans l'ADN.

Une réaction de résolution *in vitro* utilise comme substrat une molécule pseudo-co-intégrée. (Le substrat a été construit en clonant *res* de telle sorte que sont trouvées des répétitions directes à deux endroits dans le plasmide pBR322.) Le substrat doit être superenroulé ; sa résolution conduit à deux cercles enchevêtrés portant chacun un site *res*. La réaction nécessite des quantités importantes de résolvasse TnpR ; la résolution fonctionne mieux lorsque le rapport enzyme sur ADN est proche d'environ 20 : 1. Il n'y a pas besoin de facteurs hôtes (contrairement à la recombinaison spécifique de site chez *lambda*).

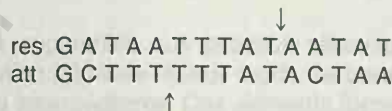
La résolution est une réaction conservative ; les liaisons sont coupées et reliées sans besoin d'un apport d'énergie. Les produits mettent en évidence une étape intermé-

diaire dans la résolution de la structure co-intégrée ; ils sont constitués de la résolvasse liée de façon covalente aux deux extrémités 5' des coupures double brin effectuées au site *res*. La coupure a lieu de façon symétrique dans une courte région palindromique conduisant à des extensions de deux bases. En agrandissant la région de croisement localisée dans le site I, on peut schématiser la réaction de coupure de la façon suivante :



La relation entre le site *res* et le site cœur *att* de *lambda*, décrite dans le Chapitre 28, entraîne une similitude avec le système de recombinaison, site spécifique du phage. Aux alentours immédiats du point de croisement, 10 des 15 pb sont homologues dans les deux sites.

Les séquences sont :



Les flèches indiquent les sites clivés dans *res* par TnpR et dans *att* par Int. Les réactions elles-mêmes sont analogues, bien que la résolution s'effectue seulement entre les sites intramoléculaires alors que la recombinaison entre les sites *att* est intermoléculaire et directionnelle (comme le montrent les différences dans les sites *att*). Une autre relation existe, mise en évidence par la séquence de la résolvasse. Cette dernière est proche de la protéine Hin qui assure l'inversion site spécifique (voir Chapitre 28). Le mécanisme de l'inversion site spécifique peut être semblable à celui de la réaction de résolution. La ressemblance existant entre les réactions de coupure et de réunion assurées par les topo-isomérases est évidente dans tous les cas.

Enfin, nous devons mentionner un effet remarquable de Tn3 appelé **immunité de transposition**. Un plasmide portant Tn3 est protégé d'insertions ultérieures provenant de tout autre Tn3 porté par une molécule différente d'ADN (bien que le Tn3 sur le plasmide soit capable lui-même de se transposer au sein du plasmide). Il y a un délai entre l'insertion de Tn3 dans un nouveau plasmide et l'établissement de l'immunité. L'effet est spécifique ; il ne concerne que Tn3 et n'empêche pas les autres transposons de s'insérer dans le plasmide. La base de cet effet, agissant en *cis* et à long terme, est encore mystérieuse.

POUR EN SAVOIR PLUS

Les différentes hypothèses développées sur les transposons peuvent être abordées grâce aux revues de **Kleckner** (*Cell* **11**, 11-23, 1977), **Calos & Miller** (*Cell* **20**, 579-595, 1980), **Kleckner** (*Ann. Rev. Genet.* **15**, 341-404, 1981), **Campbell** (*Ann. Rev. Microbiol.* **35**, 55-83, 1981) s'est intéressé à leur signification

dans l'évolution. Les mécanismes de la transposition ont fait l'objet d'une revue par **Grindley & Reed** (*Ann. Rev. Biochem.* **54**, 863-896, 1985). Deux chapitres de *Mobile Genetic Elements* (ed. Shapiro, Academic Press, New York, 1983) sont consacrés à des transposons particuliers : celui de **Heffron** sur Tn3 et celui de **Kleckner** sur Tn10.

CHAPITRE 30

LES ÉLÉMENTS MOBILES CHEZ LES EUCARYOTES

Les éléments transposables se propagent avec autant de succès chez les eucaryotes que chez les procaryotes. Le patrimoine génétique des eucaryotes comprend une variété d'éléments qui ont la capacité de se déplacer au sein du génome dans lequel ils résident vers de nouveaux sites choisis au hasard.

Il faut supposer que ces éléments possèdent les gènes nécessaires permettant de coder pour la transposition ; nous ne savons pas s'ils portent en plus des fonctions accessoires qui n'ont aucun rapport avec celle-ci. Certains de ces éléments peuvent, comme les transposons bactériens, utiliser des mécanismes qui permettent aux séquences d'ADN de se déplacer, soit directement, soit par production de copies génomiques. Comme les transposons bactériens, ils présentent à leurs extrémités de courtes séquences inversées répétées et engendrent au niveau de leur site d'insertion dans l'ADN hôte, de brèves séquences directement répétées. Ces éléments n'ont aucune existence en dehors du génome, comparables en cela aussi aux transposons bactériens. Qu'ils soient indispensables à la survie cellulaire ou qu'ils soient des parasites uniquement concernés par leur propre survie, ces éléments n'ont pas d'existence indépendante et n'engendrent pas de molécules d'ADN libres.

Un autre mode de transposition est représenté par la capacité qu'ont les rétrovirus d'insérer des copies ADN (les provirus) d'un ARN viral génomique dans les chromosomes de la cellule hôte. Certains transposons eucaryo-

tes sont apparentés aux provirus rétroviraux par leur organisation générale et leur mode de transposition via des ARN intermédiaires. Ces éléments forment la classe des **rétrotransposons**. Ils vont des rétrovirus eux-mêmes, qui sont capables d'infecter les cellules hôtes, aux séquences qui ont subi un événement de transposition via un ARN intermédiaire quoiqu'étant dépourvues par elles-mêmes de la capacité de se transposer.

Même dans les génomes où aucun transposon n'a pu être mis en évidence, il est trouvé des traces d'anciens événements de transposition sous la forme de séquences répétées directes accolées à des séquences répétitives dispersées dans le génome. Les caractéristiques de ces séquences impliquent parfois qu'il y ait eu une séquence ARN à l'origine de la séquence génomique (ADN). On pense que l'ARN a dû être converti en une copie d'ADN double brin avant d'être intégré dans le génome par un phénomène proche de la transposition.

LES ÉLÉMENTS DE CONTRÔLE DU MAÏS SONT TRANSPOSABLES

Les études génétiques du maïs, amorcées en 1940, ont mis en évidence des modifications du génome pendant la division cellulaire somatique. Ces modifications sont apportées par les **éléments de contrôle**, connus pour leur capacité à se déplacer d'un site à un autre. Mis en

évidence à l'origine par Mac Clintock qui a simplement utilisé des analyses génétiques, ces éléments sont maintenant considérés comme éléments transposables.

L'insertion d'un élément de contrôle peut modifier l'activité des gènes adjacents. Des délétions, des duplications, des inversions et des translocations peuvent avoir lieu au site d'intégration de ces éléments. La cassure chromosomique est une conséquence habituelle de la présence de certains éléments. Une caractéristique propre au maïs est que la fréquence et la chronologie des événements associés à la présence d'éléments de contrôle sont soumises à une régulation pendant le développement, respectant peut-être une spécificité tissulaire.

Le génome du maïs contient plusieurs familles d'éléments de contrôle. Le nombre, les types et les localisations de ces éléments sont caractéristiques de chaque souche de maïs. Les membres de chaque famille peuvent être divisés en deux classes.

Les **éléments autonomes** ont les propriétés d'excision et de transposition ; leur insertion donne lieu à des allèles instables. Décrits tout d'abord comme allèles « aptes à muter » (mutable en anglais), la terminologie a gardé la lettre *m* pour les désigner. La perte de l'élément de contrôle lui-même, ou de sa capacité à se transposer

entraîne la conversion d'un allèle apte à muter en un allèle stable.

Les **éléments non autonomes** sont stables et deviennent instables *uniquement* lorsqu'un élément autonome de la même famille est présent ailleurs dans le génome. Quand il est complété en *trans* par un élément autonome, l'élément non autonome peut présenter les activités habituellement associées aux éléments autonomes. Les éléments non autonomes dérivent peut-être d'éléments autonomes ayant perdu les fonctions agissant en *trans*, nécessaires à la transposition.

Les éléments autonomes subissent des « changements d'état », transmissibles, mais qui sont des altérations instables de leurs propriétés. Ces changements peuvent influencer la fréquence et la chronologie de la transposition. Certaines modifications semblent entraîner la perte d'autonomie des éléments autonomes. Des altérations agissant en *cis* peuvent avoir lieu dans les éléments non autonomes ; les éléments autonomes ne peuvent alors plus agir sur ceux-ci ; donc, un élément non autonome peut devenir définitivement stable parce que sa transposition ne peut plus être activée.

Les familles des éléments de contrôle ont été définies d'après les interactions existant entre les éléments autonomes et non autonomes. Un élément non autonome

Figure 30.1

Chaque famille d'éléments de contrôle possède des membres autonomes et non autonomes.

Les éléments autonomes sont capables de se transposer et possèdent également d'autres activités (par exemple, ils peuvent influencer l'expression génique). Les éléments non autonomes sont défectifs au moins pour la transposition. A l'origine, les noms donnés à chaque paire d'éléments autonomes et non autonomes étaient classés par famille. Il peut y avoir plus d'un élément dans chaque classification ; par exemple, *Spm-s* et *Spm-w* sont des versions différentes de *Spm*.

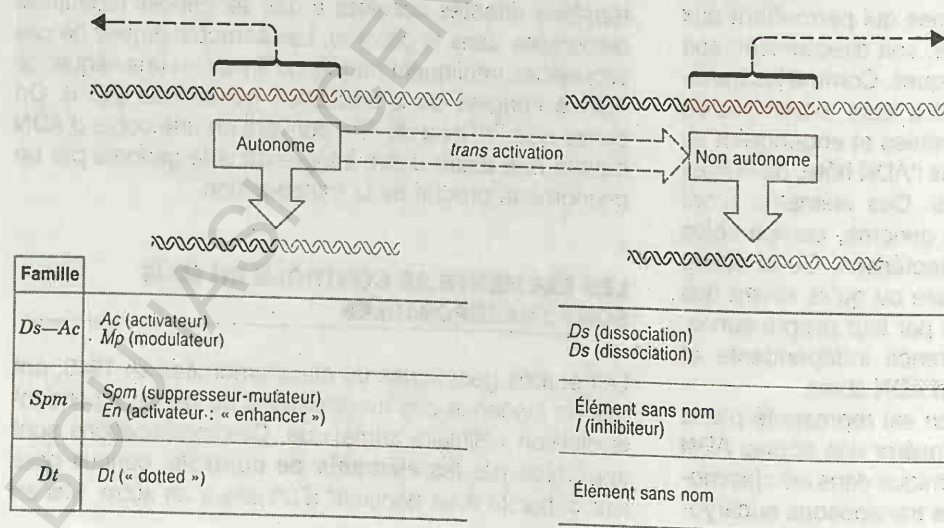


Tableau 30.1
Les transposons du maïs ont le même type d'organisation que les transposons bactériens.

Élément	Copies /Génome	Longueur	Cadres delecture	Répétitions inversées	Répétitions cibles
Ac	?	4 563 pb	3	11 pb	8 pb
Mu1	~ 5	1 367 pb	4	213 pb	9 pb
Spm (En)	?	8 287 pb	1	13 pb	3 pb

peut être activé en *trans* seulement par certains éléments autonomes ; ils sont considérés comme membres d'une même famille. Quelques familles d'éléments de contrôle, *Ac-Ds*, *Spm* et *Dt*, sont résumées dans la **Figure 30.1**. Plusieurs transposons du maïs sont maintenant caractérisés au niveau moléculaire. Ils partagent la même organisation générale — séquences répétées inversées aux extrémités et séquences directement répétées dans l'ADN hôte — mais leur taille et leur capacité codante varient. Le **Tableau 30.1** résume les propriétés de quelques transposons.

Deux caractéristiques du maïs ont permis de suivre le devenir d'événements de transposition. Les éléments de contrôle s'intègrent souvent près de gènes entraînant une modification détectable, non létale, du phénotype. Et, parce que le maïs se développe selon un mode clonal, l'apparition de la transposition et sa chronologie peuvent être visualisées comme il l'est décrit dans le diagramme de la **Figure 30.2**.

La nature de l'événement n'a pas d'importance : il peut s'agir d'une insertion, d'une excision ou d'une cassure chromosomique. En revanche, est important l'événement qui a pour effet de modifier l'expression d'un allèle chez un hétérozygote. Les descendants d'une cellule touchée par un événement dépendant d'un élément présentent un nouveau phénotype alors que les descendants d'une cellule non affectée par cet événement gardent le phénotype d'origine.

Les descendants issus de la mitose d'une cellule conservent la même localisation et donnent naissance à un même **secteur** tissulaire. Un changement phénotypique pendant le développement somatique est appelé **diversification**. Il se manifeste par l'apparition d'un secteur ayant le nouveau phénotype au sein du tissu présentant le phénotype d'origine. La taille du secteur dépend du nombre de divisions ayant eu lieu dans la lignée cellulaire qui lui a donné naissance ; par consé-

quent, la taille du secteur présentant un nouveau phénotype est déterminée par le moment où intervient le changement génotypique. Plus précoce est son apparition dans la lignée cellulaire, plus grand est le nombre de descendants et donc plus grande est la taille du secteur affecté dans le tissu mature.

Ds PEUT TRANPOSER OU ENTRAÎNER UNE CASSURE CHROMOSOMIQUE

Les éléments de contrôle peuvent avoir des fonctions multiples. Par exemple, le nom donné aux membres de la famille autonome *Spm* indique bien leurs doubles propriétés : supprimeur-mutateur. Supprimeur car pouvant inhiber l'expression d'un gène de structure au locus déjà occupé par un élément non autonome. Mutateur pour leur aptitude à se transposer de telle sorte que le gène retrouve son activité normale.

Ces deux propriétés sont portées par le même élément mais elles peuvent être distinguées par les mutants *sp⁺m⁻* qui ont perdu l'une d'entre elles. Puisque l'aptitude d'un élément non autonome à supprimer les fonctions adjacentes est contrôlée en *trans* par un élément autonome, l'effet ne peut résulter simplement de l'insertion de l'élément non autonome mais doit être la conséquence d'une régulation spécifique.

La présence de l'élément non autonome *Ds* a plusieurs conséquences sur le locus adjacent. Activé par un élément autonome *Ac*, *Ds* peut soit se transposer vers un nouveau site, soit entraîner une cassure chromosomique. La transposition est accompagnée par l'excision à partir du site donneur, qui a pour effet de convertir l'allèle instable en un allèle stable (qui n'est pas nécessairement l'allèle sauvage). Une relation inverse entre réversion vers la stabilité et apparition de cassure chromosomique

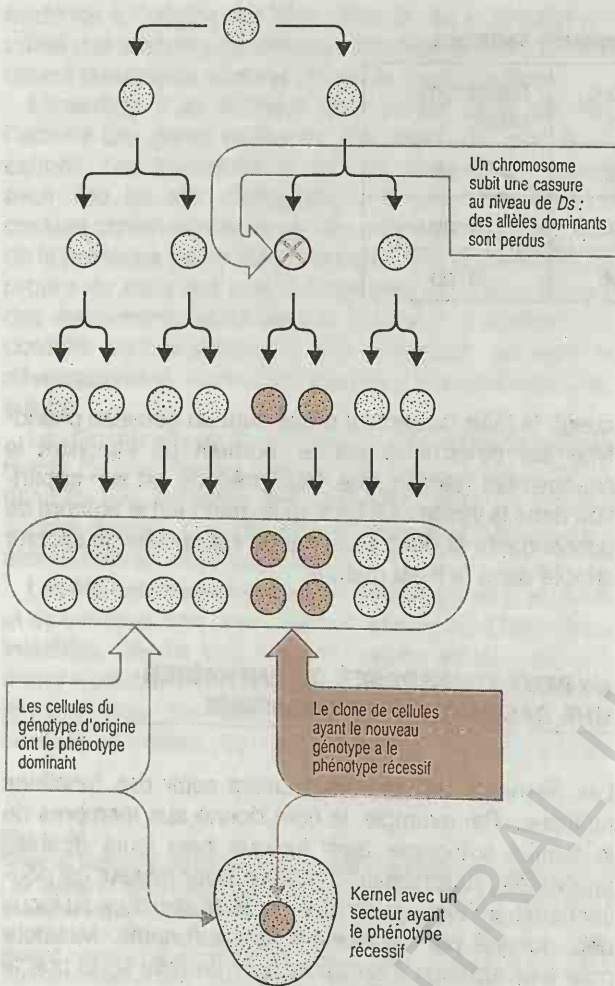


Figure 30.2

L'analyse clonale permet l'identification d'un groupe cellulaire, issu d'un seul précurseur, chez lequel un événement provoqué par la transposition a modifié le phénotype. Le nombre de cellules indique à quel moment au cours du développement cet événement a eu lieu. La localisation des cellules peut indiquer la spécificité tissulaire du processus.

Plusieurs types d'événements peuvent conduire à la perte d'allèles dominants ; la Figure 30.4 montre un exemple dans lequel la cassure entraîne la perte d'allèles dominants, permettant l'expression du phénotype récessif.

implique que la transposition et la cassure sont les deux alternatives possibles lorsque *Ds* est activé.

Les insertions des éléments *Ds* et *Ac* ont été suivies sur deux loci, *Sh* et *Wx*. D'après les changements intervenus dans les loci mutants, les séquences de plusieurs variants *Ac* et *Ds* sont connues.

Le locus *Shrunken* (*Sh*) code pour la sucrose synthase, une enzyme impliquée dans la biosynthèse de l'amidon. L'absence de l'enzyme entraîne le phénotype *shrunken* kernel qui donne son nom au locus. Les loci *Sh-m* chez lesquels *Ds* est présent, possèdent une carte de restriction de leur région 5' différente, conséquence de l'insertion d'ADN exogène associée parfois à la duplication de l'ADN du locus. La modification observée dans la séquence est différente pour chaque allèle *sh-m*.

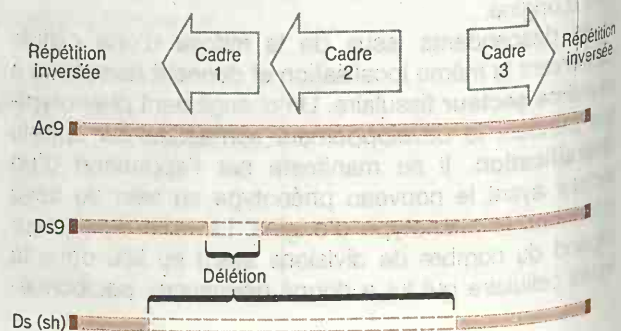
Le locus *Waxy* (*Wx*) détermine le contenu en amylose du pollen et du tissu endospermique. Il code pour une enzyme, l'UDP-glucose amidon transférase qui est liée aux grains d'amidon. Des insertions de *Ac* et *Ds* ont été trouvées. La comparaison des séquences de ces insertions avec celles trouvées au locus *Sh* permet d'établir les structures de *Ac* et de *Ds*. Celles-ci sont représentées Figure 30.3.

La longueur de l'élément autonome *Ac* est de 4 563 pb. Il possède trois principaux cadres de lecture ouverte. Les cadres 1 et 2, lus dans la même direction, semblent coder pour des protéines ainsi qu'on peut le déduire de l'enchaînement de leurs triplets. La capacité codante du cadre 3, lu dans la direction opposée, est plus incertaine. Ces régions codantes couvrent presque la totalité de la séquence de *Ac*. L'élément possède à ses extrémités une séquence de 11 pb répétée inversée. Dans l'ADN hôte, une séquence de 8 pb est dupliquée au site d'insertion.

Les éléments *Ds* ont une longueur et des séquences variables mais ils sont proches des éléments *Ac*. Ils se terminent par une même séquence 11 pb inversée répétée. L'élément *Ds9* présente une délétion de 194 pb ; l'élément *Ds6* a une longueur de 2 kb représentant 1 kb à partir de chaque extrémité de *Ac*. Un élément double

Figure 30.3

L'élément *Ac* possède deux cadres de lecture ouverte ; les éléments *Ds* présentent des délétions internes.



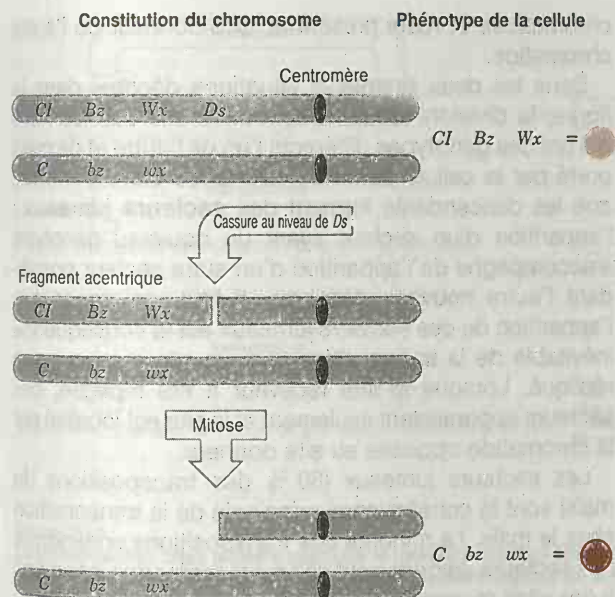


Figure 30.4
Une cassure au niveau de l'élément de contrôle entraîne la perte d'un fragment acentrique ; si le fragment porte des marqueurs dominants d'un hétérozygote, sa perte change le phénotype.

Les effets marqueurs dominants, *C*-, *Bz*, *Wx*, peuvent être visualisés par la couleur des cellules ou par une coloration appropriée. (*C*-I est un régulateur qui inhibe la formation de pigment et conduit à un aleurone incolore ; *Bz* glycosyle les pigments anthocyanides ; *Wx* est responsable de la formation de l'amylase, colorée en bleu avec les I_2 -KI.)

Ds, de structure complexe, possède deux séquences *Ds*6, l'une insérée en sens inverse par rapport à l'autre.

Les éléments non autonomes sont donc dépourvus de séquences internes mais possèdent les séquences inversées répétées terminales (et probablement d'autres caractéristiques). Les éléments non autonomes dérivent des éléments autonomes par des délétions (ou par d'autres modifications) qui inactivent la transposase agissant en trans mais qui laissent intacts les sites (y compris les extrémités) sur lesquels agit la transposase.

A l'origine, l'élément *DS* a été identifié parce qu'il pouvait, après activation par *Ac*, servir de site pour une cassure chromosomique. Les conséquences d'un tel événement sont décrites Figure 30.4. Considérons un hétérozygote chez lequel *Ds* est inséré dans un locus chromosomique situé entre le centromère et une série de marqueurs dominants. Le chromosome homologue est dépourvu de *Ds* et porte des marqueurs récessifs (*C*, *bz*, *wx*).

La cassure au niveau de *Ds* engendre un fragment acentrique qui porte les marqueurs dominants. A la mitose, ce fragment est perdu puisqu'il lui manque le

centromère. Par suite, les cellules filles possèdent uniquement les marqueurs récessifs portés par le chromosome intact. Les conséquences d'une telle situation sont décrites dans la Figure 30.2.

La cassure au niveau de *Ds* entraîne la formation de deux chromosomes inhabituels qui sont engendrés par la liaison des extrémités cassées des produits de réplication. L'un est un fragment acentrique en forme de U, constitué par l'union des chromatides sœurs de la région distale à *Ds* (à gauche, comme il l'est dessiné Figure 30.4). Le second est un chromosome dicentrique en forme de U qui comprend les chromatides sœurs de la région proximale à *Ds* (à droite, sur la figure). Cette dernière structure conduit au cycle classique de cassure-fusion-pontage représenté dans la Figure 30.5.

Considérons le devenir du chromosome dicentrique lors de sa ségrégation sur le fuseau mitotique. Chacun de ses deux centromères est attiré vers un pôle opposé. La tension casse le chromosome en un site au hasard, localisé entre les deux centromères. D'après l'exemple décrit dans la figure, la cassure a lieu entre les loci *A* et *B*, entraînant une duplication de *A* chez l'un des chromosomes fils et une délétion chez l'autre. Si *A* est un marqueur dominant, les cellules portant la duplication vont garder le phénotype *A*, alors que les cellules possédant la délétion auront le phénotype récessif *a*.

Le cycle de cassure-fusion-pontage se maintient sur plusieurs générations permettant ainsi la poursuite des modifications génétiques chez les descendants. Par exemple, intéressons-nous au chromosome portant la délétion en *A*. Au cours du cycle suivant, une cassure peut avoir lieu entre *B* et *C* conduisant à deux types de descendants, l'un avec une duplication de *B* et l'autre avec une délétion. Les pertes successives de marqueurs dominants peuvent être révélées par l'apparition de sous-secteurs au sein de secteurs.

LA TRANSPPOSITION DE *Ds* EST LIÉE À LA RÉPLICATION

La transposition de *Ds* est accompagnée par sa disparition du site donneur d'origine. Les résultats d'une analyse clonale suggèrent que la transposition de *Ds* intervient presque toujours après la réplication de l'élément donneur. Ces caractéristiques sont proches de celles observées lors de la transposition de l'élément bactérien *Tn10* (voir Chapitre 29). L'information la plus importante sur le déplacement de *Ds* est fournie par les transpositions qui ont lieu d'un site donneur à un site receveur localisés sur un même chromosome.

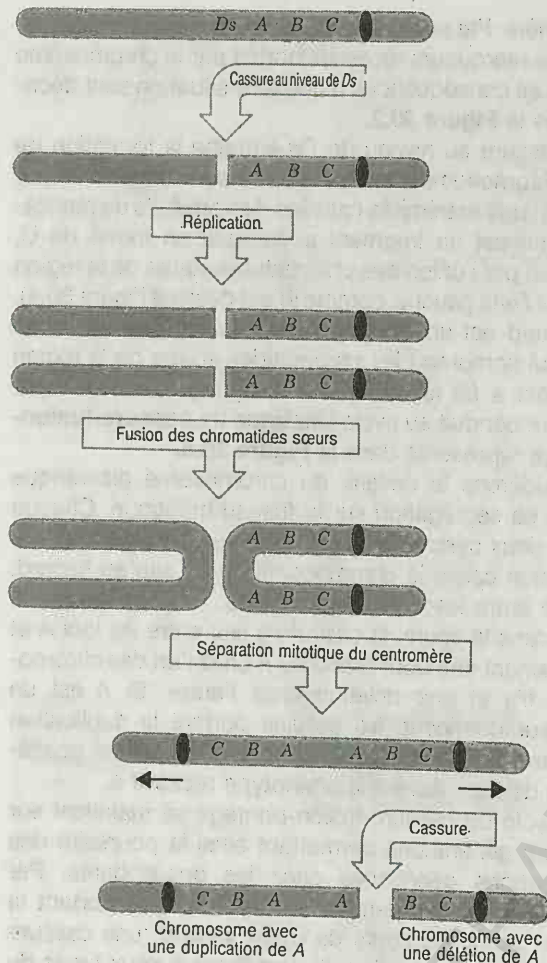


Figure 30.5

Activé par *Ac*, *Ds* fournit un site permettant d'initier le cycle fusion-pontage-cassure de la chromatide. L'analyse clonale permet de suivre les produits du cycle.

Selon que le site receveur est répliqué ou non, deux issues de transposition peuvent être distinguées. Elles sont schématisées en haut et au centre de la Figure 30.6. Si le site receveur n'est pas répliqué avant la transposition, l'élément se répliquera à son nouveau site. Il est donc présent au site receveur sur les deux chromosomes fils. Si le site receveur est préalablement répliqué, seule la chromatide recevant « physiquement » l'élément présentera un changement au locus receveur. Dans l'autre cas, l'élément est perdu au site donneur de l'une des

chromatides et reste présent au site donneur de l'autre chromatide.

Dans les deux premières situations décrites dans la figure, la division cellulaire conduit à des cellules filles qui ont des génotypes différents l'un de l'autre et de celui porté par la cellule parentale. L'analyse clonale montre que les descendants forment des **secteurs jumeaux** : l'apparition d'un secteur ayant un nouveau génotype s'accompagne de l'apparition d'un autre secteur possédant l'autre nouveau génotype. Il faut remarquer que l'apparition de ces secteurs jumeaux est la conséquence inévitable de la transposition vers un site receveur non répliqué. Lorsque le site receveur a été répliqué, ces secteurs apparaissent seulement si le site est localisé sur la chromatide opposée au site donneur.

Les secteurs jumeaux (80 % des transpositions du maïs) sont la conséquence principale de la transposition chez le maïs. La minorité des transpositions engendrant des secteurs uniques peut soit représenter des insertions à des sites receveurs situés sur la même chromatide que les sites donneurs, soit résulter des transpositions qui ne sont pas reliées à la réplication (ce qui est montré sur la ligne inférieure de la figure), soit enfin être artéfactuelle, du fait de la perte de l'un des produits de transposition (par exemple, si une cellule s'engage dans un autre lignage). Même s'il est bien clair que la transposition est habituellement associée à la réplication du site donneur, nous ne savons pas si cette relation est obligatoire.

Qu'arrive-t-il au site donneur ? Les réarrangements trouvés aux sites qui ont perdu les éléments de contrôle peuvent être expliqués par les mécanismes de réparation de cassure, secondaire à l'excision d'un élément de contrôle.

LE CYCLE DE VIE DES RÉTROVIRUS IMPLIQUE DES ÉVÉNEMENTS ANALOGUES À LA TRANSPOSITION

Les **rétrovirus** possèdent des génomes formés d'ARN simple brin qui se répliquent via un ADN double brin intermédiaire. Le cycle de vie du virus implique une étape pendant laquelle l'ADN double brin s'insère dans le génome hôte par transposition, engendrant de courtes séquences directement répétées d'ADN.

Cette réaction signifie bien plus que la seule perpétuation du virus. Certaines de ces conséquences sont les suivantes :

- Les séquences rétrovirales intégrées restent dans le

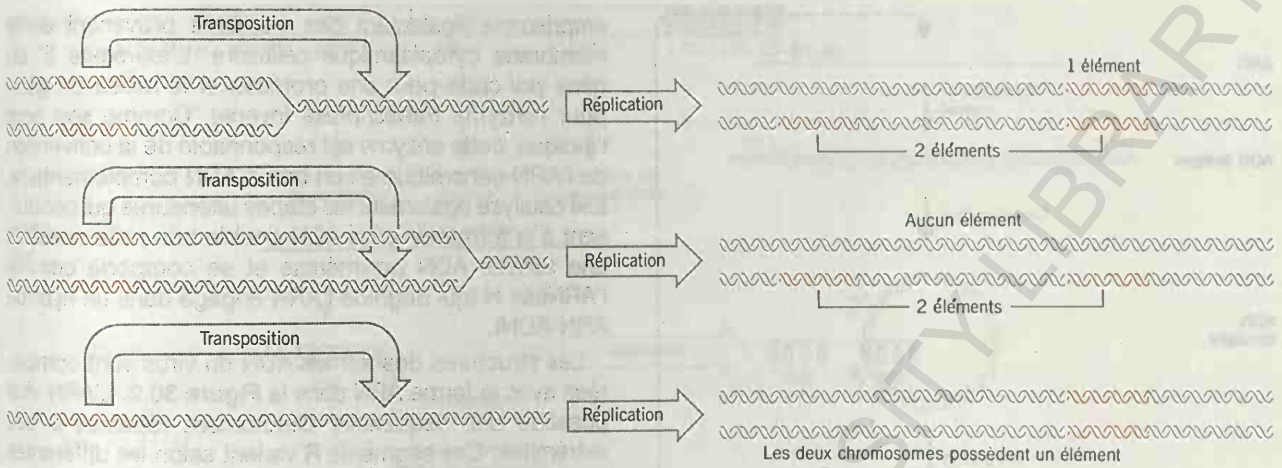


Figure 30.6

La transposition chez le maïs implique le mouvement d'un élément à partir d'un site donneur après qu'il ait été répliqué ; la nature des produits formés dépend de la répllication ou non-répllication des sites donneur et receveur avant la transposition.

En haut : Si le site receveur n'a pas été répliqué, l'élément est dupliqué après la transposition de telle sorte que chacun des chromosomes fils acquiert un élément. L'un des chromosomes perd un élément et en garde un, au site receveur ; l'autre chromosome porte à la fois des éléments aux sites donneur et receveur.

Au centre : Si le site receveur a été répliqué, seul l'un des chromosomes fils gagne un élément à ce même site. Habituellement, un élément se déplace d'un site donneur localisé sur une chromatide jusqu'au site receveur de l'autre chromatide, de telle sorte que le génotype des deux chromatides est modifié. (Si l'élément se déplace d'un site donneur à un site receveur au sein d'une même chromatide, le résultat génétique ne pourra être distingué de celui représenté sur le schéma supérieur.)

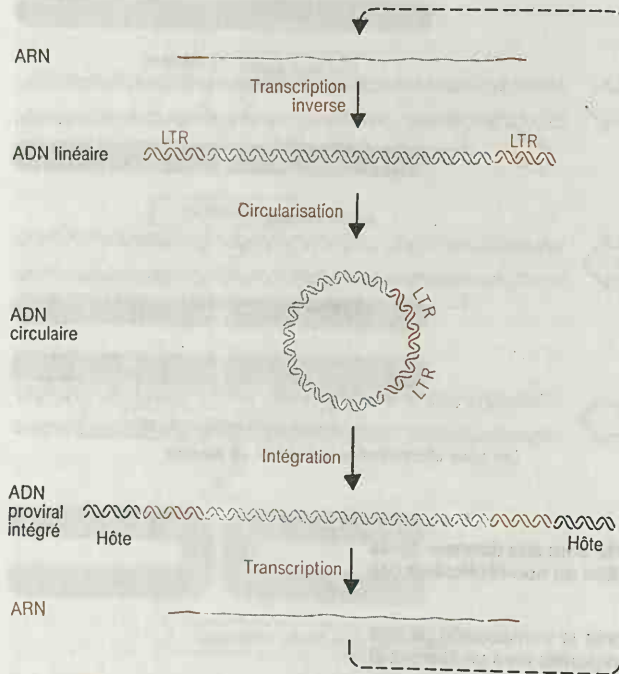
En bas : Si la transposition a lieu avant la répllication du site donneur, toute la descendance portera l'élément au site receveur.

génomique cellulaire sous la forme de **provirus endogènes**. Comme les bactériophages lysogènes, ils se comportent comme une partie du matériel génétique de l'organisme.

- Des séquences cellulaires se recombinent avec le rétrovirus et, par la suite, se transposent avec lui ; ces séquences peuvent modifier les propriétés d'une cellule infectée par ce virus recombinant.
- Des séquences autres que rétrovirales peuvent être occasionnellement « capturées » lors de ce processus et se trouver insérées sous la forme de séquences double brin dans de nouveaux sites génomiques.
- Il existe des transposons proches des rétrovirus (et ayant probablement partagé une évolution commune).

L'enzyme responsable de la copie de l'ARN en ADN est la **transcriptase inverse**. Nous en avons déjà parlé au Chapitre 16, comme d'un outil précieux dans la technologie de l'ADN recombinant. La transcriptase inverse est contenue, avec le génome, dans la particule virale. Cette enzyme convertit l'ARN en un ADN double brin linéaire dans le cytoplasme de la cellule infectée.

Le cycle rétroviral est schématisé **Figure 30.7**. L'ADN linéaire atteint le noyau (par un mécanisme inconnu) où il est converti en un ADN circulaire. Une copie d'ADN ou plus peut s'intégrer dans le génome hôte. L'**ADN proviral** intégré est transcrit par la machinerie de l'hôte en ARNs viraux qui servent à la fois d'ARNm et de génomes pour l'intégration dans les virions. *L'intégration est une étape normale du cycle et est nécessaire à la transcription.*

**Figure 30.7**

Le cycle rétroviral s'effectue grâce à la transcription inverse du génome ARN en ADN double brin qui s'insère ensuite dans le génome hôte, de façon à être transcrit en ARN.

Deux copies d'ARN génomiques sont encapsidées dans un virion ; chaque particule virale est donc diploïde. Lorsqu'une cellule est infectée par deux virus différents mais proches, il peut y avoir formation de particules virales hétérozygotes possédant un génome de chaque type. La diploïdie est peut-être importante dans le sens où elle permet au virus d'acquérir des séquences cellulaires.

Une séquence rétrovirale type contient trois « gènes », ce terme représentant ici chacune des régions codantes qui conduit à la formation de plusieurs protéines grâce à des réactions de maturation. L'ordre de ces gènes est *gag-pol-env*, comme cela est décrit **Figure 30.8**. L'ARN viral est traduit comme un messenger normal — il porte une coiffe à son extrémité 5' ou « cap », et son extrémité 3' est polyadénylée — en la polyprotéine *gag*. Par suppression ou par changement du cadre de lecture après le codon de terminaison de *gag*, l'ARN est également traduit en une protéine *gag-pol*. L'épissage engendre un messenger, subgénomique, plus court qui est traduit en la polyprotéine *env*, précurseur de deux protéines.

Le gène *gag* conduit à la formation des éléments protéiques du noyau nucléoprotéique du virion. Le gène *env* code pour les protéines de l'enveloppe virale laquelle

emprisonne également des éléments provenant de la membrane cytoplasmique cellulaire. L'extrémité 5' du gène *pol* code pour une protéase et le milieu du gène pour l'enzyme transcriptase inverse. Comme son nom l'indique, cette enzyme est responsable de la conversion de l'ARN génomique en un brin d'ADN complémentaire. Elle catalyse également les étapes ultérieures qui conduisent à la formation d'un ADN double brin ; elle possède une activité ADN polymérase et se comporte comme l'ARNase H (qui dégrade l'ARN engagé dans un hybride ARN-ADN).

Les structures des formes ADN du virus sont comparées avec la forme ARN dans la **Figure 30.9**. L'ARN viral possède des séquences directement répétées à ses extrémités. Ces segments **R** varient selon les différentes souches virales de 10 à 80 nucléotides. Après le segment R de l'extrémité 5', se trouve la région **U5** de 80-100 nucléotides, dont le nom indique qu'elle est spécifique de l'extrémité 5' (U pour « unique » en anglais). Précédant le segment R de l'extrémité 3', est localisé le segment **U3** de 170-250 nucléotides, spécifique de cette partie 3'.

Les extrémités de l'ADN linéaire possèdent des séquences supplémentaires. Un segment U3 est ajouté à l'extrémité 5' de même qu'un segment U5 à l'extrémité 3'. Par conséquent, chaque extrémité de l'ADN porte la même séquence **U5-R-U3** ; celle-ci est appelée **LTR** (Long Terminal Repeat : longue répétition terminale). Elle est formée par la réaction décrite **Figure 30.10** pendant laquelle la *transcriptase inverse* change de matrice emportant avec elle l'ADN en cours de synthèse. La Figure décrit la formation d'un LTR ; un processus semblable est nécessaire à la formation du LTR de l'autre extrémité.

Les LTRs à chaque extrémité du virus sont identiques. L'extrémité 3' de U5 est constituée d'une courte séquence répétée inversée par rapport à celle présente dans l'extrémité 5' de U3, de telle sorte que le LTR lui-même se termine par de courtes séquences répétées inversées. L'ADN proviral intégré ressemble à un transposon : la séquence provirale se termine par des séquences répétées inversées et elle est flanquée de courtes séquences directement répétées dans l'ADN hôte.

La forme rétrovirale impliquée dans la réaction d'intégration est une molécule circulaire qui contient deux copies en tandem du LTR (voir **Figure 30.7**). « L'intégrase » reconnaît la séquence formée par la jonction des deux LTR ; elle reconnaît apparemment les régions externes des séquences répétées inversées, s'y positionne et effectue le clivage nécessaire à l'intégration. La fonction « intégrase » est apportée par la séquence codée par l'extrémité 3' du gène *pol*.

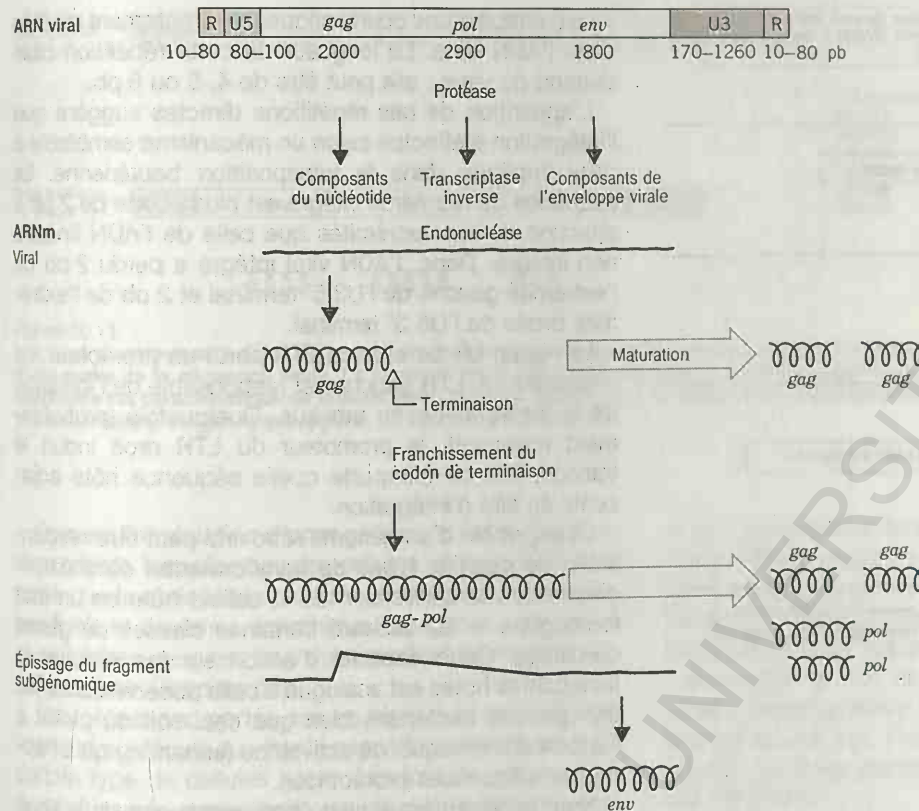


Figure 30.8
Les « gènes » rétroviraux sont exprimés sous forme de protéines chevauchantes qui sont maturées en produits individuels.

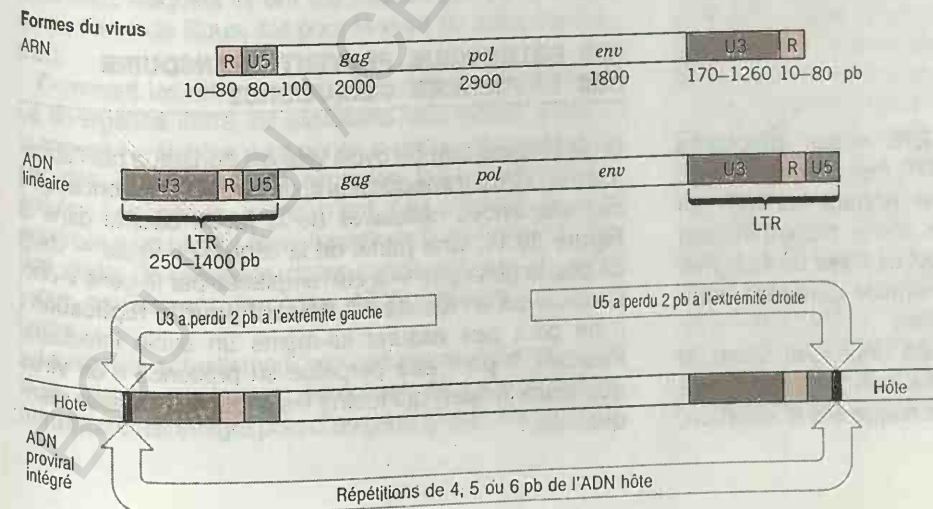


Figure 30.9
L'ARN rétroviral se termine dans les répétitions directes (R), l'ADN linéaire libre s'étend jusqu'aux LTRs et le provirus jusqu'aux LTRs raccourcis chacun de deux bases.

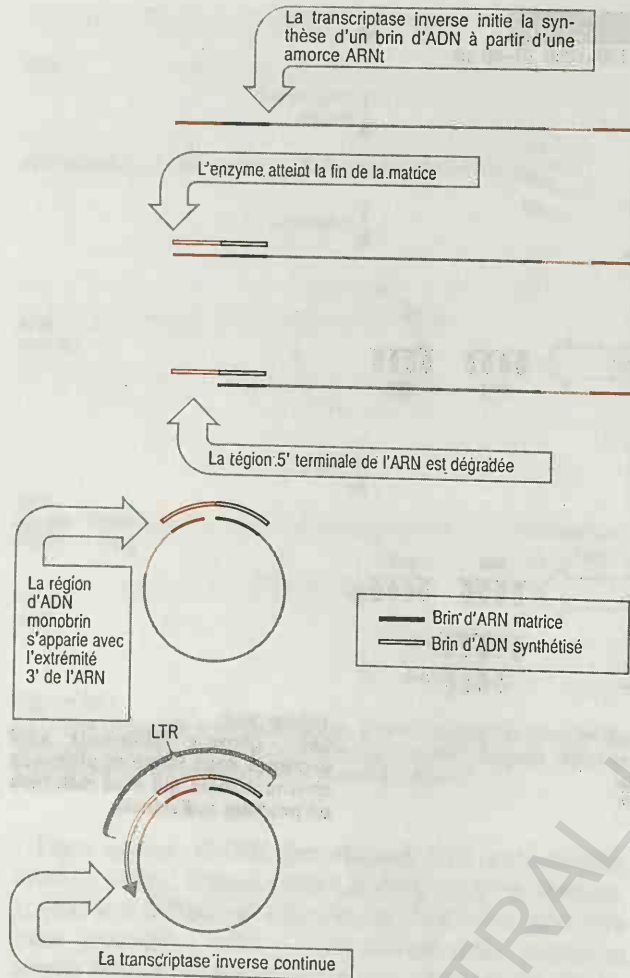


Figure 30.10

Le LTR est engendré pendant la transcription inverse grâce aux changements de matrices. La nouvelle matrice peut être l'autre extrémité de la même molécule (comme le montre le schéma), ou une autre molécule.

On trouve également des ADNs viraux circulaires contenant une seule copie du LTR ; ces formes sont en fait plus nombreuses que celles portant les LTR en tandem. Elles peuvent provenir d'une recombinaison homologue entre les LTRs de provirus libres ou intégrés. Les LTRs isolés, trouvés dans certains génomes, sont peut-être des vestiges de l'excision.

L'ADN viral s'intègre au hasard dans des sites du génome hôte. Une cellule infectée peut posséder entre une et dix copies du provirus. A chaque site d'insertion,

il y a création d'une courte séquence directement répétée dans l'ADN hôte. La longueur de cette répétition cible dépend du virus ; elle peut être de 4, 5 ou 6 pb.

L'apparition de ces répétitions directes suggère que l'intégration s'effectue selon un mécanisme semblable à celui impliqué dans la transposition bactérienne. La séquence du rétrovirus intégré est plus courte de 2 pb à chacune de ses extrémités que celle de l'ADN linéaire non intégré. Donc, l'ADN viral intégré a perdu 2 pb de l'extrémité gauche de l'U3 5' terminal et 2 pb de l'extrémité droite de l'U5 3' terminal.

La région U3 de chaque LTR porte un promoteur. Le promoteur du LTR gauche est responsable de l'initiation de la transcription du provirus. Quelquefois (probablement rarement), le promoteur du LTR droit induit la transcription de n'importe quelle séquence hôte adjacente au site d'intégration.

L'intégration d'un génome rétroviral peut être responsable de certains types de transformation cellulaire — c'est-à-dire la conversion de la cellule hôte en un état tumorigène — en activant certaines classes de gènes cellulaires. Cette capacité d'action sur les séquences flanquantes hôtes est analogue à celle observée chez les transposons bactériens bien que ceci soit dû plutôt à l'apport d'une séquence activatrice (enhancer) qu'à l'apport de séquences promotrices.

Nous avons envisagé jusqu'à présent le cycle infectieux rétroviral dans lequel l'intégration est nécessaire à la production d'autres copies d'ARN. Toutefois, lorsqu'un ADN viral s'intègre dans une cellule de la lignée germinale, il est transmis comme virus endogène de l'organisme. Les virus endogènes ne s'expriment habituellement pas mais ils peuvent parfois être activés par des événements externes, par exemple une infection par un autre virus.

LES RÉTROVIRUS PEUVENT TRANSDUIRE DES SÉQUENCES CELLULAIRES

Un fait intéressant du cycle viral est constitué par l'apparition de **virus transducteurs**, des variants qui ont acquis des séquences cellulaires de la façon décrite dans la **Figure 30.11**. Une partie de la séquence virale — dans ce cas, le gène *env* — a été remplacée par le gène *v-onc*. Le virus qui en résulte est **défectif pour la réplication** ; il ne peut pas assurer lui-même un cycle infectieux. Pourtant, il peut être perpétué en présence d'un **virus auxiliaire** (helper) qui fournit les séquences virales manquantes.

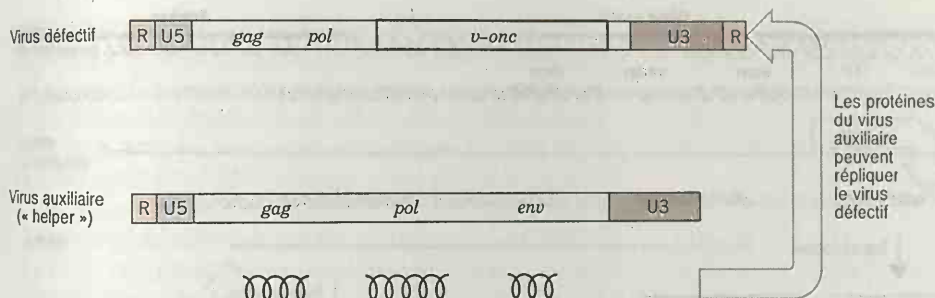


Figure 30.11

Les virus transformants défectifs pour la réplication possèdent une séquence cellulaire à la place d'une partie de la séquence virale. La longueur de la séquence cellulaire et de la séquence virale substituée est caractéristique de chaque virus. Le virus déficient peut se répliquer en présence d'un virus auxiliaire (« helper ») qui apporte les fonctions manquantes.

Onc est l'abréviation d'**oncogénèse**, c'est-à-dire de la capacité de **transformer** des cellules en culture de telle sorte que la croissance cellulaire échappe aux contrôles habituels et devienne anarchique. Les gènes *onc*, viraux et cellulaires, peuvent tous deux rendre tumorigènes des cellules (voir Chapitre 35).

Les gènes *onc* portés par les rétrovirus sont appelés *v-onc* et confèrent au virus la capacité de transformer un certain type de cellules hôtes. Les loci du génome, qui portent les séquences homologues, sont appelés *c-onc*. A leur état normal, ils interviennent dans certaines fonctions cellulaires (souvent de façon inconnue). Une mutation peut occasionnellement activer un gène *c-onc* lequel entraîne une tumorigénisation cellulaire. Un gène *c-onc* peut être aussi activé par l'intégration d'un rétrovirus à sa proximité, comme il l'est décrit dans le sous-chapitre précédent.

Les noms donnés aux gènes *onc* dérivent de ceux des virus chez lesquels ils ont été trouvés (*src* pour le virus du sarcome de Rous, *fes* pour le virus du sarcome félin, etc.).

Comment les rétrovirus acquièrent-ils les gènes *onc* ? La divergence entre les structures des gènes *c-onc* et *v-onc* est révélatrice. Les gènes *c-onc* sont interrompus par des introns et sont exprimés faiblement dans la cellule. Les gènes *v-onc* n'ont pas d'intron et sont exprimés à un niveau élevé à partir de l'unité de transcription virale. De telles structures suggèrent que les gènes *v-onc* proviennent de copies ARN, épissées, de gènes *c-onc*.

Le modèle actuellement proposé pour la formation des virus transformants est représenté Figure 30.12. Un rétrovirus s'est intégré près d'un gène *c-onc*. Une délétion

a lieu, entraînant la fusion du provirus au gène *c-onc* ; la transcription engendre ensuite un ARN de fusion qui contient les séquences virales à l'une de ses extrémités et les séquences cellulaires à l'autre. Les introns cellulaires sont ensuite excisés. L'ARN possède les signaux nécessaires à son encapsidation dans le virion ; des virions seront produits si la cellule porte une autre copie intacte du provirus. Par conséquent, certaines particules virales diploïdes pourront contenir un ARN de fusion et un ARN viral.

Une recombinaison entre ces séquences peut conduire à la formation du génome transformant dans lequel les séquences répétées virales sont présentes à *chaque* extrémité. (La recombinaison a lieu très fréquemment pendant le cycle infectieux rétroviral ; elle utilise certainement des intermédiaires ADN ou des réactions additionnelles de transfert de brins. Rien n'est connu du degré d'homologie requis entre les brins d'ADN ; il faut supposer que la réaction non homologue, qui a lieu entre le génome viral et la partie cellulaire de l'ARN de fusion, s'effectue grâce à des mécanismes identiques à ceux responsables de la recombinaison virale.)

Les caractéristiques communes à l'ensemble de la classe rétrovirale suggèrent qu'elle dérive peut-être d'un ancêtre unique. Les éléments IS primordiaux ont pu se placer de part et d'autre d'un gène codant pour une acide nucléique-polymérase constituant une unité ayant la structure *LTR-pol-LTR*. Celle-ci a pu évoluer vers le virus infectieux par l'acquisition de fonctions plus sophistiquées permettant l'utilisation de substrats ADN et ARN, ainsi que l'encapsidation de l'ARN en virion. D'autres fonctions, comme les gènes transformants, ont pu être incorporées plus tard. (Il n'y a pas de raison de penser

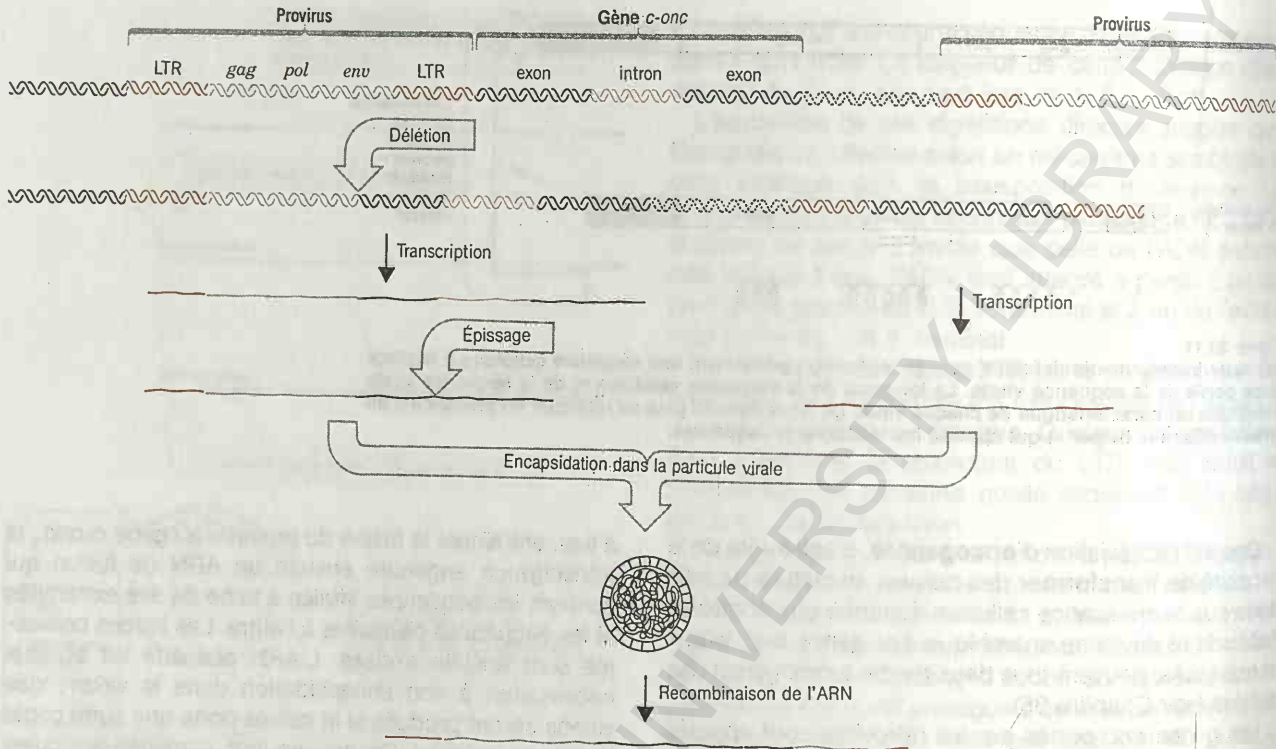


Figure 30.12

Les virus transformants défectifs pour la réplication peuvent être formés via l'intégration et la délétion d'un génome viral qui conduit à un transcrit cellulaire et viral fusionné pouvant être encapsidé avec un ARN génomique normal. La recombinaison non homologue est nécessaire à la formation du génome transformant défectif pour la réplication.

que le mécanisme impliqué dans l'acquisition de fonctions cellulaires ne concerne que les gènes *onc* ; mais les virus portant ces gènes présentent peut-être un avantage sélectif du fait de leur pouvoir stimulateur de la croissance cellulaire.)

DES TRANPOSITIONS ARN-DÉPENDANTES ONT PU AVOIR LIEU DANS LA CELLULE

Certains pseudogènes présentent des caractéristiques internes et externes suggérant qu'ils proviennent de séquences d'ARN, bien que dans ce cas, on ne puisse que spéculer sur les processus ayant abouti à la synthèse de la copie d'ADN. Les caractéristiques d'un tel pseudogène sont comparées à celles du gène d'origine et de l'ARNm dans la Figure 30.13. Cette figure regroupe toutes les caractéristiques significatives mais seules cer-

taines d'entre elles sont observées dans chaque cas particulier.

Le pseudogène peut commencer au point correspondant à l'extrémité 5' de l'ARN, ce que l'on attend si l'ADN provient de l'ARN. Plusieurs pseudogènes sont constitués de séquences exoniques liées de façon précise ; nous savons qu'il n'y a aucun mécanisme de reconnaissance des introns dans l'ADN. Ceci renforce donc l'hypothèse d'une étape nécessitant un ARN intermédiaire. Le pseudogène se termine parfois par une courte répétition de paires de bases A-T qui dérivent probablement de la queue de poly(A) de l'ARN. Une courte séquence directement répétée peut être trouvée de part et d'autre du pseudogène ; elle provient probablement d'un événement semblable à la transposition.

Les pseudogènes maturés (rétro-pseudogènes ou processed genes en anglais) se localisent à des sites qui n'ont aucun rapport avec ceux de leur origine supposée ; ceci renforce l'hypothèse selon laquelle ces pseudogènes

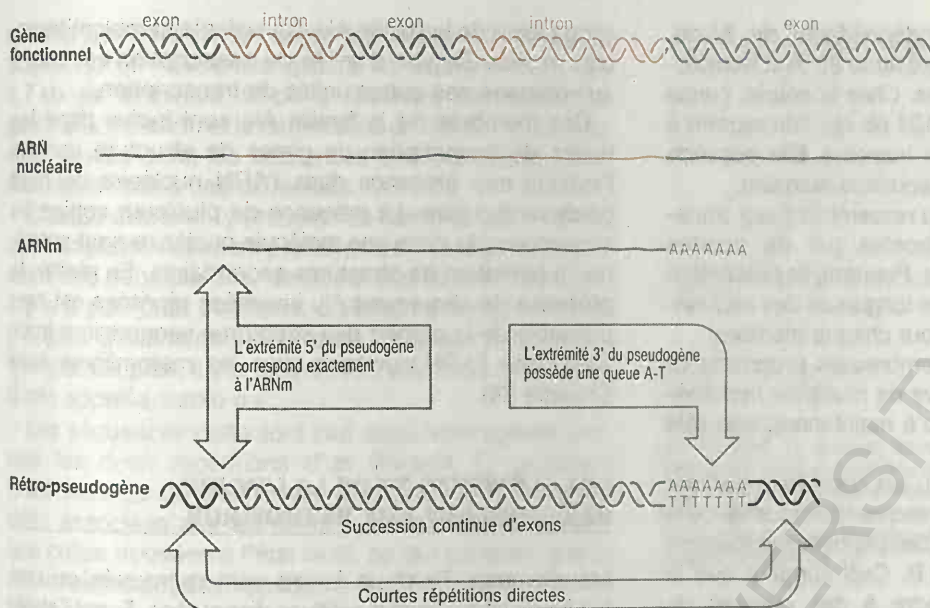


Figure 30.13
Des pseudogènes proviennent peut-être de la transcription inverse de l'ARN qui conduit à l'intégration d'un ADN double brin dans le génome.

ont été formés par un autre mécanisme que celui impliqué pour les pseudogènes trouvés à proximité des groupes de gènes actifs.

Les pseudogènes maturés diffèrent à bien des égards des rétrovirus : les séquences supposées s'être transposées ne portent aucune information pouvant être utilisée pour la transposition (ou pour l'étape précédente de transcription inverse de l'ARN). Un rétrovirus est-il responsable de ce processus ? Ou est-ce la conséquence d'un système cellulaire aberrant ? Peut-être est-ce de façon fortuite que les extrémités de la séquence transposée ressemblent aux séquences trouvées aux extrémités d'un transposon.

Des événements de transposition surviennent-ils actuellement dans ces génomes ou ne constate-t-on que les empreintes d'événements anciens ? Il faut noter que pour « survivre », les transpositions doivent avoir lieu dans la lignée germinale ; des événements semblables doivent probablement survenir dans les cellules somatiques mais ils ne sont pas transmis aux générations suivantes.

LA FAMILLE Alu

Dans certains génomes de mammifères, une proportion non négligeable des séquences dispersées moyenne-

ment répétitives comprend des membres d'une même famille. Leur courte longueur et leur degré élevé de répétition les rendent comparables à l'ADN de séquence monotone excepté le fait que les membres de cette famille sont dispersés dans le génome au lieu d'être regroupés en tandems multiples.

Dans le génome humain, une grande partie de l'ADN moyennement répétitif existe sous forme de séquences de 300 pb qui sont dispersées au sein de l'ADN non répétitif. L'ADN double brin correspondant à la séquence moyennement répétée peut être isolé par renaturation à un C_{ot} intermédiaire suivie de la dégradation de l'ADN non répétitif restant non apparié. Au moins la moitié du matériel double brin renaturé est clivée par l'enzyme de restriction Alu I, en un site situé toutes les 170 pb.

Les séquences clivées sont les membres d'une même famille appelée **Alu** d'après la façon dont elle a été identifiée. Il y a à peu près 300 000 membres dans le génome haploïde (soit en moyenne un membre tous les 6 kb d'ADN). Les séquences Alu sont très répandues. Une famille proche est trouvée chez la souris (où les 50 000 membres forment la famille B1), chez le hamster chinois (où elle est appelée famille Alu-équivalente) et chez d'autres mammifères.

Les membres de la famille Alu sont apparentés plutôt qu'identiques. La famille humaine semble provenir d'une duplication en tandem de 130 pb possédant dans sa

moitié droite une séquence indépendante de 31 pb. Chaque séquence individuelle présente 87 % d'homologie avec la séquence consensus. Chez la souris, l'unité répétée B1 a une longueur de 131 pb qui correspond à la taille du monomère de l'unité humaine. Elle possède 70-80 % d'homologie avec la séquence humaine.

Les membres de la famille Alu ressemblent aux transposons ; ils sont également bordés par de courtes séquences directement répétées. Pourtant, ils présentent une caractéristique curieuse : les longueurs des séquences répétées sont différentes pour chaque membre.

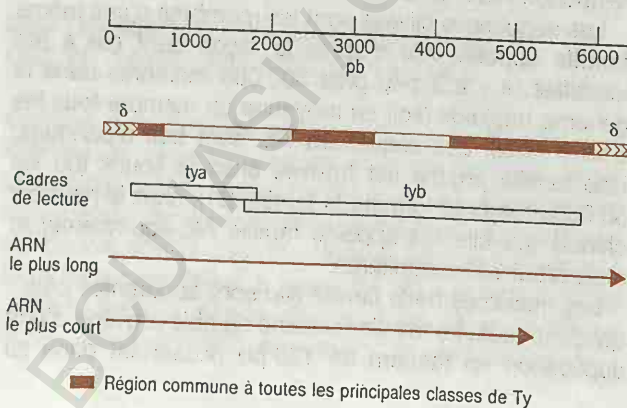
La famille Alu possède de nombreuses propriétés et son caractère ubiquitaire soulève de multiples hypothèses quant à sa fonction ; jusqu'à maintenant, son rôle réel n'a pas été déterminé.

Une partie de la séquence Alu est constituée par une séquence de 14 pb qui est presque identique à celle trouvée à l'origine de réplication des papovavirus (comme SV40) et du virus de l'hépatite B. Ceci suggère que la famille Alu pourrait correspondre à des origines de réplication du génome eucaryote. Le grand nombre d'éléments de cette famille va à l'encontre de cette hypothèse puisqu'il y a environ 10 fois plus de séquences Alu qu'il n'y a d'origines de réplication supposées dans une cellule eucaryote (voir Chapitre 13).

Une autre propriété est observée : au moins certains membres de la famille sont transcrits *in vitro* en un petit ARN par l'ARN polymérase III (responsable de la transcription des petits ARNs nucléaires, des ARNs t, et de l'ARN 5S). Chez le hamster chinois, certains membres

Figure 30.14

Les éléments Ty se terminent par de courtes séquences directement répétées et ils sont transcrits en deux ARNs chevauchants. Ils ont deux cadres de lecture ouverte dont les séquences sont proches de celles des gènes rétroviraux *gag* et *pol*.



(sinon tous) de la famille Alu-équivalente semblent transcrits *in vivo*. Ce type d'unités de transcription est trouvé au voisinage des autres unités de transcription.

Des membres de la famille Alu sont inclus dans les unités de transcription de gènes de structure comme l'indique leur présence dans l'ARN nucléaire de haut poids moléculaire. La présence de plusieurs copies de séquences Alu dans une molécule nucléaire peut entraîner la formation de structures secondaires. En réalité, la présence de séquences Alu inversées répétées est responsable de la plupart des structures secondaires trouvées dans l'ARN nucléaire chez les mammifères (voir Chapitre 24).

LES ÉLÉMENTS TY DE LA LEVURE RESSEMBLENT AUX RÉTROVIRUS

Les éléments Ty de la levure sont regroupés en une famille de séquences répétitives dispersées. Ty est l'abréviation pour « transposon yeast » ; l'existence d'une fonction de transposition est suggérée par l'observation que des éléments Ty sont trouvés à différents sites dans différentes souches. La transposition laisse une empreinte caractéristique : de chaque côté de l'élément Ty intégré, 5 pb de l'ADN hôte sont répétées. La fréquence de transposition de l'élément Ty paraît moins élevée que celle des transposons bactériens, environ 10^{-7} - 10^{-8} .

Les éléments Ty divergent beaucoup entre eux. La plupart des éléments Ty se répartissent dans l'une des deux principales classes appelées Ty1 et Ty917. Les relations existant entre ces dernières sont décrites dans la Figure 30.14. Chaque élément a une longueur de 6,3 kb ; les 330 dernières paires de bases constituent à chaque extrémité des répétitions directes appelées δ . Tous les éléments présentent ces répétitions, puis une région longue du côté gauche, une autre région centrale et enfin une petite région adjacente aux répétitions δ de l'extrémité droite. Les deux régions de substitution principale ne possèdent aucune homologie.

La séquence de l'élément Ty a deux cadres de lecture ouverte, exprimés dans la même direction mais lus dans des phases différentes se chevauchant de 13 acides aminés. D'après sa séquence, *tya* pourrait coder pour une protéine se liant à l'ADN. La séquence de *tyb* présente des homologies avec les séquences codant pour la transcriptase inverse.

L'élément Ty est transcrit en deux ARN polyadénylés qui constituent plus de 5 % des ARNm totaux de la cellule haploïde de levure. L'initiation de la transcription est située à 95 pb de l'extrémité droite. L'un des ARNm

s'étend sur 5 kb se terminant à l'intérieur de la région commune de l'extrémité gauche et l'autre s'étend sur 5,7 kb se terminant 40 pb à l'intérieur de l'extrémité gauche. L'initiation et la terminaison de cet ARN plus long sont situées dans les séquences δ et il possède donc des répétitions à ces extrémités.

Chaque élément Ty présente de nombreuses différences comparé à l'élément prototype de leur classe, notamment des substitutions, des délétions et des insertions. Il y a à peu près 30 copies d'éléments de type Ty1 et 6 copies d'éléments Ty917 dans un génome de levure typique. De plus, il y a plus de 100 éléments δ indépendants appelés « solo δ ».

Les séquences delta sont tout aussi hétérogènes bien que les deux répétitions d'un élément Ty semblent identiques ou du moins très proches. Les séquences delta associées aux éléments Ty sont plus conservées que celles trouvées à l'état isolé, ce qui suppose que la reconnaissance des répétitions est impliquée dans la transposition.

Les éléments Ty fournissent des régions mobiles d'homologie qui sont les cibles des recombinaisons effectuées par les systèmes de l'hôte. Un tel événement entraîne habituellement l'altération du chromosome soit par la création d'une délétion ou d'une inversion (par recombinaison entre deux éléments Ty sur un chromosome), soit par des modifications encore plus importantes lorsque la recombinaison s'effectue entre deux éléments Ty situés sur des chromosomes différents.

La recombinaison entre les éléments Ty a lieu apparemment par à-coups ; lorsqu'un tel événement est détecté, plus grande est la probabilité d'en observer d'autres. La conversion génique est une des interactions existant entre les éléments Ty ; elle a lieu entre des éléments ayant des localisations différentes de telle sorte qu'un élément est « remplacé » par un autre.

L'excision des éléments Ty peut s'effectuer par recombinaison homologue entre les séquences delta directement répétées. Le grand nombre d'éléments « solo delta » est peut-être la trace laissée par de tels événements. Une excision de ce type peut entraîner la reversion d'une mutation causée par l'insertion d'un élément Ty ; le degré de la reversion dépend des séquences delta restantes après l'excision.

Paradoxalement, bien que les éléments delta possèdent la même séquence, le promoteur est trouvé dans la séquence delta droite et la terminaison de la transcription dans la séquence delta gauche. (Une caractéristique semblable est observée chez d'autres éléments transposables dont les rétrovirus.)

Les éléments Ty sont des rétrotransposons classiques,

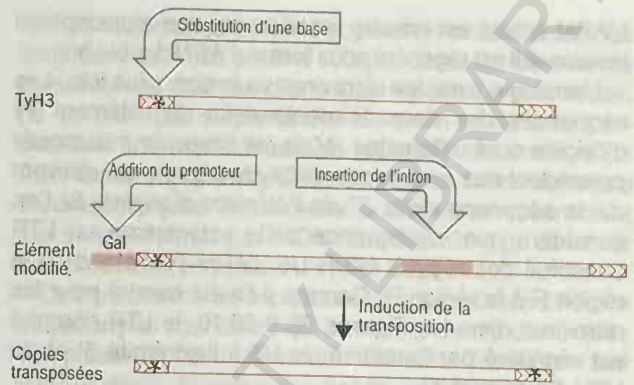


Figure 30.15

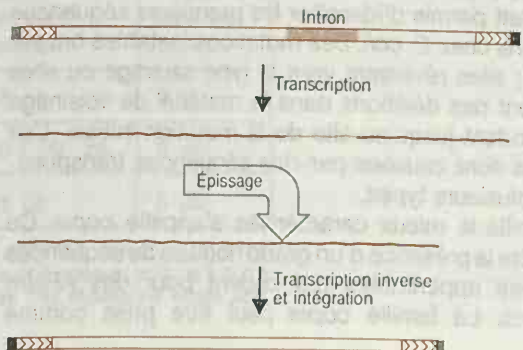
Un élément Ty unique, dans lequel un intron a été inséré, se transpose pour donner des copies ne contenant pas l'intron. Les copies possèdent des répétitions terminales identiques, provenant de l'une des extrémités de l'élément Ty d'origine. La fréquence de la transposition est fonction de l'activité du promoteur qui contrôle la transcription de Ty.

se transposant via un ARN intermédiaire. Ceci a pu être mis en évidence grâce au protocole ingénieux décrit dans la **Figure 13.15**. Un intron est inséré dans un élément formant ainsi une séquence Ty particulière. Cette séquence est placée sous le contrôle du promoteur GAL dans un plasmide et ce dernier est introduit dans les cellules de levure. La transposition entraîne l'apparition de nombreuses copies du transposon dans le génome de levure, mais elles ont toutes perdu l'intron.

Nous savons qu'il n'y a qu'un seul moyen d'éliminer les introns : la réaction d'excision-épissage de l'ARN. On peut alors proposer le modèle suivant, décrit dans la **Figure 30.16**. L'élément Ty est transcrit en un ARN qui est reconnu par le système assurant l'excision-épissage.

Figure 30.16

Les éléments Ty se transposent suivant le même processus que celui utilisé par les rétrovirus.



L'ARN épissé est ensuite reconnu par une transcriptase inverse et il est régénéré sous forme d'ADN double brin.

L'analogie avec les rétrovirus va encore plus loin. Les séquences des deux éléments delta de l'élément Ty d'origine sont différentes. *Mais les éléments transposés possèdent des séquences delta identiques, qui dérivent de la séquence delta 5' de l'élément d'origine.* Si l'on considère que la séquence delta est comme un LTR constitué des régions U3-R-U5, l'ARN Ty s'étend de la région R à la région R. Comme il l'a été montré pour les rétrovirus dans les Figures 30.8-30.10, le LTR complet est régénéré par l'addition de U5 à l'extrémité 3' et de U3 à l'extrémité 5'.

Des gènes portés par l'élément Ty contrôlent la transposition. Le promoteur GAL qui dirige la transcription de l'élément Ty modifié est inductible : il est activé par l'addition de galactose. L'induction du promoteur a deux effets : d'une part, elle est nécessaire pour activer la transposition de l'élément modifié, et d'autre part, elle augmente la fréquence de transposition des éléments Ty sur le chromosome de levure.

Le cadre de lecture *tyb* code pour la transcriptase inverse qui ressemble à la protéine rétrovirale. L'activité transcriptase inverse sédimente comme si elle était associée à une particule ; des particules ressemblant à des virus peuvent être observées dans des cellules où il y a eu induction de la transposition. Ceci augmente encore l'analogie entre le transposon Ty et le rétrovirus. Mais il est pourtant peu probable que l'élément Ty donne naissance à des particules infectieuses ; il se comporte un peu comme on pourrait l'attendre d'un rétrovirus qui aurait perdu son gène *env* et qui serait donc incapable d'encapsider normalement son génome.

IL Y A DE NOMBREUX ÉLÉMENTS TRANSPOSABLES CHEZ *D. MELANOGASTER*

L'existence d'éléments transposables chez *D. melanogaster* a pu être déduite d'observations semblables à celles ayant permis d'identifier les premières séquences d'insertions chez *E. coli*. Des mutations instables ont été trouvées ; elles réversent vers le type sauvage ou elles engendrent des délétions dans le matériel de voisinage qui s'étendent jusqu'au site de la mutation initiale. Ces mutations sont causées par des séquences transposables de plusieurs types.

La famille la mieux caractérisée s'appelle *copia*. Ce nom reflète la présence d'un grand nombre de séquences étroitement apparentées qui codent pour des ARNm abondants. La famille *copia* peut être prise comme

paradigme pour plusieurs autres classes d'éléments dont les séquences sont éloignées mais dont les structures et les comportements sont proches.

Le nombre de copies de l'élément *copia* dépend de la souche de mouche ; il est habituellement de 20 à 60. Les membres de la famille sont très dispersés. Lorsque les génomes de mouches de différentes souches sont comparés par analyse de restriction, les fragments hybridants avec une sonde spécifique de *copia* varient selon les souches (tout en étant chevauchants). L'hybridation *in situ* sur les chromosomes polytènes montre que des séquences *copia* sont présentes au niveau de séries de sites différents quoique là encore chevauchants. Cette variabilité de la distribution de *copia* est illustrée par le fait qu'aucun site particulier n'est occupé par *copia* dans chacune des souches individuelles examinées.

Ces différences sont apparues au cours de l'évolution. Il y a peu de changements entre des souches qui ont divergé récemment (il y a à peu près 40 ans, c'est-à-dire depuis leur utilisation en laboratoire). A l'heure actuelle, nous n'avons aucune idée de la vitesse de changement. Toutefois, la culture des cellules peut donner des indications sur des événements passés inaperçus. Le nombre d'éléments *copia* par génome augmente sensiblement : de plus de 2 à 3 fois. Les éléments supplémentaires sont des éléments *copia* intégrés dans des nouveaux sites. L'adaptation à la culture, par un mécanisme inconnu, augmente transitoirement le taux de transposition qui atteint de 10^{-4} à 10^{-3} événements par génération.

La structure de l'élément *copia* est décrite dans la Figure 31.17 et est détaillée dans le Tableau 30.2. Il a une longueur de 5 000 pb et possède des séquences terminales directement répétées identiques de 276 pb. Ces dernières présentent à leurs extrémités des répétitions inversées. Une répétition directe de 5 pb d'ADN hôte est engendrée au site d'insertion. Les membres de la famille *copia* divergent peu entre eux : moins de 5 % ; les variants présentent souvent de petites délétions. Toutes ces caractéristiques sont communes à toutes les familles de type *copia* bien que leurs membres puissent diverger davantage, jusqu'à environ 20 % pour l'élément 297.

Le caractère identique pour chaque élément *copia* des deux répétitions directes implique soit qu'elles interagissent pour permettre des corrections, soit qu'elles ont été formées pendant la transposition, à partir d'une des séquences directement répétées d'un élément fondateur. Nous venons de décrire cette propriété pour les éléments Ty de la levure ; par déduction, la relation entre rétrovirus et élément *copia* semble évidente. *Copia* possède un

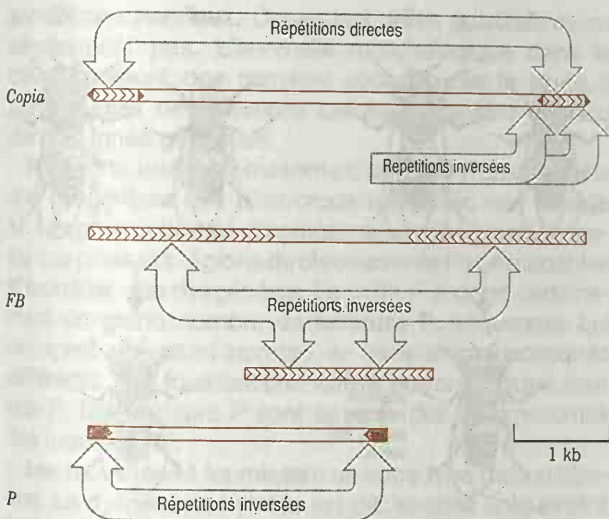


Figure 30.17 Les trois types d'éléments transposables de *D. melanogaster* ont des structures différentes.

cadre unique de lecture ininterrompu qui pourrait coder pour des régions *gag* et *pol* semblables aux séquences rétrovirales.

Les éléments *copia* demeurent toujours intacts dans le génome ; des copies isolées des répétitions terminales n'ont pas été trouvées (bien que l'on puisse s'attendre à cette éventualité si la recombinaison entraîne la délétion du matériel central). En revanche, des éléments *copia* sont trouvés parfois sous forme circulaire libre, ce qui n'existe pas pour les transposons bactériens. Les deux

formes majoritaires de cette population circulaire ont des longueurs de 5 000 pb et 4 700 pb. Cette dernière forme présente une délétion d'une des répétitions terminales, ce qui n'est pas surprenant si elle provient de la recombinaison entre les séquences répétées terminales d'un ADN linéaire (pré-intégré) ou entre des éléments *copia* chromosomiques. Ces molécules circulaires ressemblent aux cercles formés par l'ADN rétroviral. On a pu observer des particules contenant un ARN *copia*.

La transcription de *copia* conduit à la synthèse d'ARNm polyadénylés abondants qui correspondent à des copies complètes et incomplètes de l'élément. Ces ARNm ont une extrémité 5' commune car ils sont initiés au milieu d'une des répétitions terminales. L'ARNm de 2 kb correspondant à une copie partielle code pour une protéine de 51 000 daltons dont on ne connaît pas la fonction. Puisque les séquences directement répétées sont identiques, il est possible qu'elles fonctionnent, comme les LTR rétroviraux, en tant que promoteurs : l'une dirigeant la transcription de l'élément afin de produire des ARN *copia*, l'autre agissant sur la transcription de gènes adjacents situés en aval de l'insertion.

Bien que nous manquions d'informations sur le mode de transposition de *copia*, les nombreuses ressemblances avec l'organisation rétrovirale mènent à la conclusion inéluctable que *copia* a une origine proche de celle des rétrovirus. Il est difficile d'établir le nombre de fonctions virales possédées par *copia*. Bien sûr, nous savons qu'il se transpose mais (de même que les éléments Ty), il n'y a aucune preuve de sa capacité infectieuse.

Les membres d'une autre famille d'éléments transposables appelée *FB* (une abréviation pour « foldback » ou repli en français) possèdent des séquences répétées

Tableau 30.2 Chez *D. melanogaster*, les éléments transposables se répartissent selon trois classes.

Élément	Copies /génome	Longueur	Répétitions directes terminales	Répétitions inversées terminales	Répétitions directes cibles
<i>copia</i>	20-60	5 000 pb	276 pb	13 pb	9 pb
<i>FB</i>	~ 30	500-5 000 pb	aucune	250-1 250 pb	9 pb
<i>P</i>	~ 50 ou 0	500-2 900 pb	aucune	31 pb	8 pb

Les éléments ayant une organisation proche de celle de *copia* sont : 412, 297, 17,6, *mgd1*, *mgd3*, *B104*, *roo* et *gypsy*. Leurs nombres de copies varient de 10 à 100 par génome, leurs longueurs sont de 5 500 à 8 500 pb, et leurs répétitions directes de 269 à 571 pb. Les longueurs des répétitions cibles sont habituellement de 4 ou 5 pb.

inversées terminales de longueur variable. Certains éléments FB sont constitués uniquement de répétitions inversées juxtaposées ; chez d'autres, les répétitions inversées sont séparées par une région d'ADN non répétitif.

Malgré leur longueur variable, les répétitions inversées de tous les membres de la famille FB sont homologues. Cette propriété est expliquée par leur structure qui consiste en des copies en tandem d'un ADN de séquence monotone, séparées par des régions plus longues de séquences plus diverses. De l'extrémité vers la partie interne de l'élément, la longueur de l'unité de séquence monotone augmente. De 10 pb, elle passe à 20 pb pour atteindre finalement 31 pb.

Les deux copies des répétitions inversées d'un même élément FB ne sont pas identiques. Les répétitions ressemblent en première analyse à de l'ADN satellite et il serait intéressant d'étudier les rapports existant entre les répétitions des différents membres de la famille.

La structure des extrémités des éléments FB est une énigme ; ce qui leur confère leur aptitude à se transposer demeure encore inconnu. Quelquefois, deux éléments FB (différents) semblent coopérer afin de transposer un grand segment d'ADN intercalé, ce qui rappelle les mécanismes utilisés par les transposons bactériens composites (bien que la longueur d'ADN comprise entre les éléments FB puisse être beaucoup plus grande). Deux exemples de telles unités ont été trouvés ; l'une est longue de 200 kb et est appelée TE (pour « transposable element » en anglais). Il est possible que chaque séquence d'ADN bordée par des éléments FB à ses extrémités se comporte comme une telle unité.

Les éléments P, beaucoup plus petits, sont un autre type d'éléments transposables. Leur longueur maximum est de 2,9 kb. Ils se terminent par de courtes répétitions inversées et ils ne se transposent que dans certaines conditions (voir ci-dessous).

LE RÔLE DES ÉLÉMENTS TRANSPOSABLES DANS LA DYSGÉNÉSIE HYBRIDE

Il est difficile de croiser entre elles certaines souches de *D. melanogaster*. Si deux mouches de ces souches sont croisées, la descendance peut présenter des « caractères » dysgéniques, c'est-à-dire un ensemble de modifications défavorables à la propagation d'un organisme dont des mutations, des aberrations chromosomiques, une ségrégation méiotique anormale et une stérilité. L'apparition de ces altérations de façon corrélée est appelée **dysgénésie hybride**.

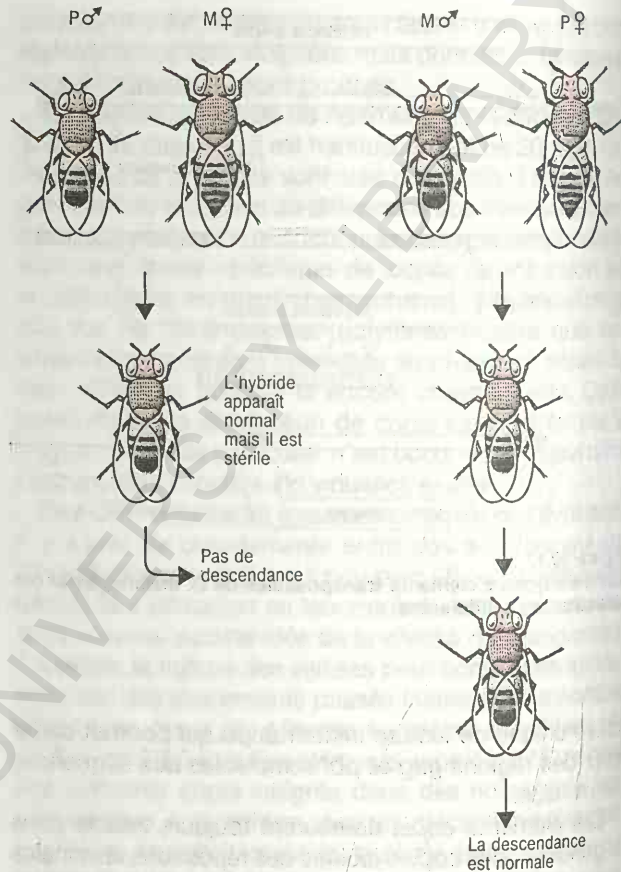


Figure 30.18

La dysgénésie hybride est asymétrique. Elle est induite par des croisements mâle P × femelle M et non par des croisements mâle M × femelle P.

Chez *D. melanogaster*, deux systèmes responsables de la dysgénésie hybride ont été identifiés. Dans le premier, les mouches sont divisées en type I (inducteur) et R (réactif). Une fertilité réduite est observée lors de croisements entre mâles I et femelles R, et non l'inverse. Dans le deuxième système, les mouches sont divisées en type P (contribution paternelle) et M (contribution maternelle). L'asymétrie du système est décrite dans la **Figure 30.18** ; un croisement entre un mâle P et une femelle M entraîne une dysgénésie et non l'inverse.

La dysgénésie apparaît surtout dans les cellules germinales. Dans les croisements mettant en jeu le système P-M, les mouches hybrides F1 possèdent des tissus

somatiques normaux. Cependant, leurs gonades ne se développent pas. L'anomalie morphologique dans le développement des gamètes apparaît dès le stade à partir duquel des divisions cellulaires rapides ont lieu dans la lignée germinale.

N'importe quel chromosome d'un mâle P peut induire une dysgénésie lors d'un croisement avec une femelle M. La construction de chromosomes recombinants montre que plusieurs régions du chromosome P sont capables d'entraîner une dysgénésie. Le mâle P a donc certainement un grand nombre de **facteurs P**, séquences qui occupent un grand nombre de sites chromosomiques différents. Ces localisations varient suivant chaque souche P. Les facteurs P sont absents des chromosomes des mouches M.

Les mouches M fournissent un autre type de contribution. La dysgénésie hybride est déclenchée uniquement lorsque les facteurs P sont mis en présence du cytoplasme transmis par la mouche femelle M. La contribution maternelle est appelée cytotype M ; les bases sur lesquelles repose cette contribution ne sont pas comprises ; mais il ne semble pas qu'il s'agisse d'un simple effet maternel, puisqu'il y a transmission sur plus d'une génération lorsqu'une femelle M est présente dans la descendance. Les rares lignées issues des croisements entre mâle P et femelle M sont converties par les facteurs P en cytotype P, c'est-à-dire en un état où les facteurs P sont inactifs.

Les événements responsables de l'induction des mutations dans la dysgénésie ont été étudiés par la cartographie de l'ADN des mutants *w* qui sont trouvés parmi les hybrides dysgéniques. Toutes les mutations consistent en l'insertion d'ADN au locus *w*. Les insertions P restent stables tant que le chromosome est dans un environnement de cytotype P, mais deviennent instables, comme le montre leur excision, lors d'une exposition au cytotype M. En plus des insertions P, les mouches dysgéniques possèdent des éléments *copia* insérés.

Les insertions P ont une longueur variable mais leurs séquences sont homologues. Les plus longues sont d'environ 2,9 kb et possèdent quatre cadres de lecture ouverte. L'élément plus court provient — apparemment assez fréquemment — d'un élément P complet ayant subi des délétions internes. Tous les éléments P possèdent des répétitions directes à leurs extrémités, comme il est décrit dans la Figure 30.17 et détaillé dans le Tableau 30.2.

Une souche P porte 30 à 50 copies du facteur P dont un tiers est de longueur complète. Les facteurs sont absents des souches M. Dans une souche P, les facteurs se comportent comme des composants inertes du géno-

me. Mais leur transposition est activée en présence du cytotype M. Un élément P actif code pour une transposase dont la synthèse est réprimée dans le cytotype P, mais activée dans le cytotype M.

Au moins certains des éléments P les plus courts ont perdu la capacité de produire la transposase mais ils peuvent être activés *in trans* par l'enzyme codée par un élément P complet. L'analyse des extrémités d'un facteur P peut indiquer si la transposition est possible. L'interaction dysgénique P-M peut aussi activer le déplacement de *copia* (ainsi que d'autres éléments transposables) ; ces insertions restent cependant stables quel que soit l'environnement cytotypique du chromosome.

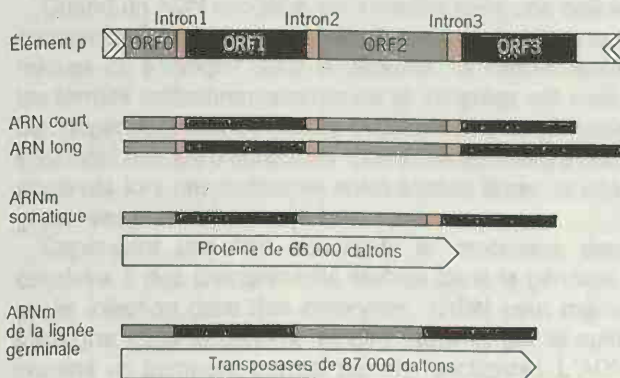
Les chromosomes des mouches hybrides dysgéniques portent plusieurs facteurs P insérés dans de nombreux sites nouveaux. Les cassures chromosomiques, caractéristiques de la dysgénésie hybride, surviennent en des points précis qui sont les sites où sont localisés des éléments P. La vitesse moyenne de transposition des éléments P sur les chromosomes M est différente chez les hybrides dysgéniques mâle et femelle mais on peut l'estimer à un événement par génération.

L'activation des éléments P est spécifique d'un tissu : elle n'a lieu que dans la lignée germinale. Mais la transcription des éléments P a lieu dans la lignée germinale et dans les tissus somatiques. *La spécificité tissulaire est conférée par un épissage différentiel.*

La Figure 30.19 décrit l'organisation et les transcrits de l'élément. Le transcrit primaire a une longueur de 2,5 ou 3 kb, cette différence reflétant probablement l'imprécision du site de terminaison. Deux épissages, l'un

Figure 30.19

L'élément P a quatre exons. Les trois premiers sont épissés ensemble lors d'une expression somatique ; et les quatre sont épissés l'un à l'autre lors de l'expression germinale.



joignant ORF0 à ORF1 et l'autre unissant ORF1 à ORF2, sont à l'origine de l'ARN détecté chez les embryons précoces. La traduction de cet ARNm conduit à une protéine de 66 000 daltons, correspondant aux trois premiers cadres de lecture ouverte.

D'après l'analyse de leurs mutations, nous savons que chacun des quatre cadres de lecture est nécessaire à l'activité de l'élément P. Il existe des sites potentiels d'épissage pouvant unir ORF2 à ORF3. Ces derniers étant mutés *in vitro* et l'élément P ainsi modifié introduit chez la mouche, la transposition est abolie. L'expérience inverse qui consiste à unir ORF2 à ORF3 en déléant l'intron conduit à un élément P qui se transpose dans tous les tissus. La protéine produite a une masse de 87 000 daltons et correspond aux quatre cadres de lecture ; elle possède l'activité transposase.

L'élément P devient actif à partir du moment où ORF3 est lié au précédent cadre de lecture. Normalement, ce processus intervient uniquement dans la lignée germinale. Sa régulation est cruciale car, sinon, la transposition a également lieu dans les cellules somatiques. Comment s'effectue la régulation de cet épissage respectant une spécificité tissulaire ? Première possibilité, un facteur présent dans la lignée germinale, peut-être une snRNP spécifique (petite ribonucléoprotéine nucléaire), reconnaît l'intron ORF2-ORF3. Dans ce cas, nous pouvons nous attendre à ce que les modifications du système d'épissage affectent aussi d'autres ARNs requis pour un développement normal. Une autre possibilité est que des modifications de l'élément P — par exemple, la formation d'une structure tertiaire particulière — affectent la sensibilité du transcrit au système d'épissage.

Les souches de *D. melanogaster* descendant de mouches capturées dans la nature il y a plus de 30 ans sont toujours M. Les souches provenant d'insectes attrapés depuis les 10 dernières années sont presque toujours P. Est-ce que cela signifie que la population de *D. melano-*

gaster sauvage a été récemment colonisée par l'élément P ? Les éléments P sont effectivement très invasifs lorsqu'ils sont introduits dans une nouvelle population ; l'élément colonisateur a pu prendre sa source chez d'autres espèces.

Parce que la dysgénésie hybride limite les croisements, elle est une étape dans la voie de la spéciation. Le croisement entre mouches provenant de deux populations géographiquement éloignées peut être rendu impossible par un phénomène de dysgénésie. Si les deux populations sont génétiquement isolées, nous aurons assisté à un phénomène de spéciation. De multiples systèmes dysgéniques, conduisant à l'impossibilité de certains croisements, peuvent ainsi être des moteurs de la spéciation.

POUR EN SAVOIR PLUS

Les éléments transposables chez les eucaryotes ont été l'objet d'une revue très approfondie dans « *Mobile Genetic Elements* » (ed. Shapiro, Academic Press, New York, 1983). Les chapitres particulièrement intéressants sont les suivants : Une analyse des éléments Ty de la levure est effectuée par Fink ; la caractérisation des transposons chez *D. melanogaster* et la cartographie du locus *white* sont décrites par Rubin ; Varmus analyse la dualité du rétrovirus, à la fois virus et élément transposable ; Fedoroff décrit les apports de la biologie moléculaire à l'étude des éléments de contrôle du maïs ; Haber résume clairement le type de croisement de la levure. Un autre résumé général concernant les transposons eucaryotes a été effectué par Finnegan (*Int. Rev. Cytol.* **93**, 281-326, 1985). Les pseudogènes maturés font l'objet d'une revue écrite par Vanin (*Ann. Rev. Genet.* **19**, 253-272, 1985). Boeke et al. ont découvert le mode de transposition utilisé par Ty (*Cell*, **40**, 491-500, 1985). Les éléments P de la *drosophile* font l'objet d'un résumé rédigé par Engels (*Ann. Rev. Genet.* **17**, 315-344, 1983) ; la base déterminant la spécificité germinale a été découverte par Laski, Rio & Rubin (*Cell*, **44**, 7-19, 1986).

Chapitre 31

MODIFICATIONS DE L'ORGANISATION DU GÉNOME

De la découverte des transpositions et des réarrangements de l'ADN découle la notion que les séquences d'ADN sont étonnamment adaptables. Des séquences peuvent être déplacées à l'intérieur du génome, modifiées ou même perdues de façon naturelle ; elle peuvent aussi être introduites expérimentalement dans les cellules.

L'existence de mécanismes naturels capables de modifier le contenu du génome semble en contradiction avec la notion générale que l'organisation du génome est relativement stable, uniquement altérée par des recombinaisons génétiques qui se produisent au sein de la lignée germinale mais jamais dans les cellules somatiques. Il est certainement vrai que des changements quantitatifs ou qualitatifs dans le génome des lignées germinales ou somatiques sont l'exception plutôt que la règle, qu'ils soient mesurés par la fréquence des événements totaux ou par la proportion du génome affecté. Cependant ils sont intéressants non seulement en eux-mêmes mais aussi par l'avantage que nous pouvons en tirer pour introduire des modifications dans le génome.

Des exemples de réarrangements ou de pertes de certaines séquences sont légions chez les eucaryotes inférieurs. Habituellement ces changements impliquent des cellules somatiques ; la lignée germinale restant inviolée. (Cependant il y a des organismes dont le cycle réplcatif comporte la perte de chromosomes entiers ou d'ensembles de chromosomes). Chez les animaux, la

réorganisation de séquences particulières est rare, bien qu'un exemple important soit représenté par le système immunitaire (voir Chapitre 32).

Au cours du développement somatique, l'altération dans les proportions relatives des composants du génome, permet aux larves d'insectes d'augmenter le nombre de copies de certains gènes. L'**amplification** occasionnelle de gènes dans des cultures de cellules de mammifères est indiquée par notre capacité à sélectionner certains variants cellulaires avec un nombre augmenté de copies de certains gènes. Initié dans le génome, l'évènement d'amplification peut créer des copies additionnelles d'un gène qui survivent sous une forme intra-ou extrachromosomique.

Quand un ADN exogène est introduit dans une cellule eucaryote, il peut engendrer des formes extrachromosomiques ou s'intégrer dans le génome. La relation entre les formes extrachromosomiques et intégrées est variable, dépendant plus du hasard et d'événements en partie imprévisibles que d'échanges réglés, ressemblant à ceux observés lors des passages entre formes libres et intégrées des plasmides bactériens.

Cependant une fois accompli, le processus peut conduire à des changements stables dans le génome ; après injection dans des embryons, l'ADN peut même s'intégrer dans le génome et être transmis par la suite comme un composé normal parfois fonctionnel. L'ADN

injecté peut être présent dans les lignées somatiques et germinales. La possibilité d'introduire des gènes particuliers fonctionnant de manière appropriée pourrait devenir une technique médicale majeure pour traiter des maladies génétiques*.

Par conséquent des manipulations considérables des séquences d'ADN sont accomplies aussi bien naturellement qu'expérimentalement. Nous commençons juste à élucider les mécanismes qui permettent à la cellule de répondre aux pressions sélectives en changeant sa batterie de séquences ou de s'adapter à l'intrusion de séquences additionnelles.

DES VARIATIONS SPÉCIFIQUES D'UN TISSU SE PRODUISENT DANS LE GÉNOME DE *DROSOPHILA*

Le contenu du génome reste inchangé dans la plupart des cellules somatiques. Cependant nous connaissons maintenant certaines situations dans lesquelles la proportion de certaines séquences est modifiée ; l'exemple le mieux caractérisé se produit au cours du développement des larves des insectes. L'ajustement prend la forme de sous ou sur-réplication de séquences particulières.

Des sous-réplications se produisent au cours du développement des tissus polytènes de la mouche. Dans ces tissus les chromosomes géants, décrits Chapitre 25, sont formés par de multiples duplications successives du lot diploïde de chromosomes (appariés).

Environ un quart du génome de *D. melanogaster* est contenu dans les régions d'hétérochromatine, qui s'agregent pour former le chromocentre (dont une grande partie est constitué d'ADN satellite). La quantité relative d'hétérochromatine est moins importante dans les cellules polytènes que dans les cellules diploïdes. L'explication en est que cette partie du génome ne se duplique pas au cours de la polyténisation de l'ADN euchromatique. La mesure du contenu en ADN satellite suggère que ces séquences ne se dupliquent pas ou peu de fois en comparaison des neuf duplications de l'ADN euchromatique des glandes salivaires.

Les gènes des ARNr sont eux aussi sous-répliqués dans les tissus polytènes. Les régions d'ADNr passent par environ 6-7 duplications (atteignant en fait le même niveau final selon que l'un ou les deux organisateurs nucléolaires sont présents. Ce type de contrôle pourrait être spécifique de l'ADNr).

Le génome de *D. melanogaster* peut donc être divisé en trois types de régions selon le contrôle de la réplication

dans les tissus polytènes. Une hypothèse naturelle est que les origines de réplication dans chaque type de séquences contrôlent le nombre d'événements d'initiation. Les origines dans l'ADN satellite ne seraient pas reconnues ; celles des ADNr s'arrêteraient de fonctionner en réponse à un rétro-contrôle par le nombre de gènes ARNr ; les origines présentent dans l'euchromatine seraient initiées jusqu'au bout.

Des amplifications différentielles de séquences codantes pour des protéines particulières sont observées chez les insectes. Dans les glandes salivaires de *Rhyncosciara* et d'autres Sciarides, des **nodules d'ADN (puffs)** sont engendrés. Leur apparence est à première vue similaire à celle des nodules qui représentent les bandes actives dans les chromosomes polytènes de Diptères mais les nodules de Sciarides contiennent des séquences d'ADN localement amplifiées et des séquences d'ARN. Les séquences amplifiées peuvent avoir subi jusqu'à 4 initiations additionnelles de la réplication accroissant d'un facteur 16 le nombre de copies.

Un aperçu des mécanismes possibles d'amplification différentielle est donné par l'état des gènes du chorion au cours du développement de *D. melanogaster*. Les protéines qui font le chorion (l'enveloppe de l'œuf) sont synthétisées et sécrétées par les cellules folliculaires polyplœides de l'ovaire. Chez les insectes, les gènes du chorion ont tendance à être groupés ; deux groupes ont été identifiés chez *D. melanogaster*. Avant leur expression dans les cellules folliculaires, les gènes du chorion localisés sur le chromosome X sont amplifiés jusqu'à 16 fois (4 duplications additionnelles) alors que les gènes du chromosome III sont amplifiés jusqu'à 60 fois (6 duplications additionnelles).

Dans chaque cas, l'amplification s'étend sur une distance d'environ 45 à 50 kb de chaque côté des gènes du chorion. Le niveau d'amplification maximum représente un plateau d'à peu près 20 kb entourant les gènes du chorion. Le degré d'amplification décroît graduellement de chaque côté du plateau.

Qu'est-ce qui est à l'origine de l'amplification ? La décroissance en gradient suggère que les régions où s'arrête l'amplification dans chaque molécule est hétérogène. Ceci peut être expliqué par le modèle illustré dans la **Figure 31.1**.

De multiples initiations de réplications bidirectionnelles se produisent à partir d'un point origine situé au centre de la région. La fourche de réplication progresse sur des distances allant de 10 à 50 kb. Ce modèle visualise la totalité de la région amplifiée comme un réplicon, présent

* N. d. T. : Pour des raisons pratiques aussi bien qu'éthiques, il paraît peu probable que le transfert de gènes dans les cellules germinales ait jamais une indication en pathologie humaine. En revanche la thérapie génétique somatique pourrait être promise à un avenir intéressant.

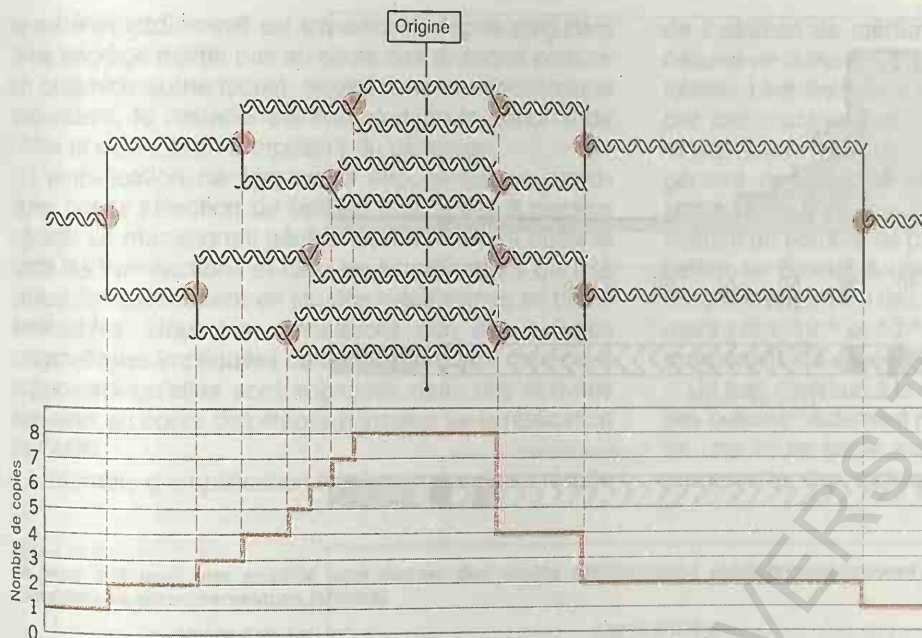


Figure 31.1

L'amplification d'une région localisée peut être accomplie par de multiples Initiations au sein d'un unique réplicon.

Chaque région répliquée est jointe aux régions bordantes non répliquées par des fourches de réplication statiques. Si les fourches arrêtent leur progression en des positions non précisément déterminées, elles engendrent un gradient comme indiqué sur la gauche. Si plusieurs fourches s'arrêtent de façon coordonnée en un site fixé, elles engendrent une brutale décroissance du niveau d'amplification comme indiqué sur la droite.

sous forme de copies multiples partiellement répliquées. (La relation entre des régions sous-répliquées ou répliquées plus fortement dans les chromosomes des glandes salivaires peut être expliquée par un modèle similaire dans lequel les jonctions entre les régions sont représentées par des fourches de réplication immobiles.)

Les régions amplifiées contiennent-elles une origine agissant en *cis*, responsable de la réplication ? Si tel est le cas, l'amplification devrait être abolie au niveau de tout le chromosome dont l'origine est mutée. Chez un hétérozygote, l'absence d'amplification n'affecterait que le chromosome muté, et la translocation d'une origine fonctionnelle guiderait l'amplification de tout réplicon dans laquelle elle serait située.

Les propriétés du mutant *ocelliless* sont conformes à ces prédictions. Cette mutation entraîne un changement complexe du phénotype ; c'est une inversion d'environ trois bandes, comprenant la région des gènes du chorion du chromosome X. Ces gènes s'étendent sur 3kb à gauche de la fin de l'inversion.

L'inversion entraîne un changement majeur du profil d'amplification. Sur la Figure 31.2, les régions amplifiées

du type sauvage et du mutant *ocelliless* sont comparées. Sur la région amplifiée d'origine, les 40 kb situés à gauche de l'inversion ne sont pas amplifiés du tout. Les 50 kb de l'autre côté sont amplifiés mais à un degré réduit atteignant seulement la moitié du maximum habituel. (L'origine de la réduction du niveau d'amplification est inconnue.) L'amplification affecte maintenant d'autres régions comme le montre le nombre augmenté de copies de séquences situées à droite, qui dans le type sauvage ne sont pas amplifiées.

Ces résultats suggèrent qu'une origine capable de gouverner une amplification bidirectionnelle réside à proximité des gènes du chorion. L'extension de la région amplifiée est identique que l'origine réside en sa localisation habituelle dans le type sauvage ou quelle ait été transférée en une nouvelle position par l'inversion *ocelliless*.

L'origine au niveau de laquelle l'amplification débute est clairement de nature plus complexe qu'une origine normale de réplication. L'origine d'amplification doit répondre avec une spécificité tissulaire et de façon dépendante du temps, à des facteurs agissant en *trans*.

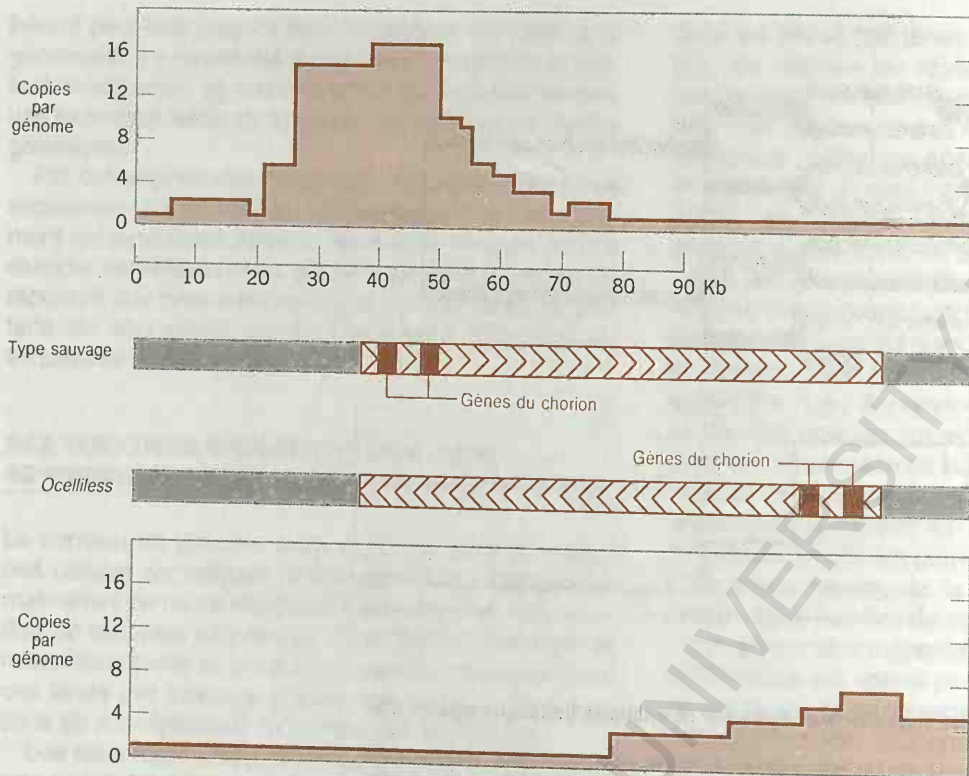


Figure 31.2

L'inversion *ocelliless* change la région d'ADN qui va être amplifiée.

L'extrémité gauche de l'inversion s'étend sur 3 kb à gauche des gènes du chorion ; l'extrémité droite n'est pas connue, mais pour les besoins de l'illustration, l'inversion est supposée s'étendre sur 85 kb.

Elle induit de nombreux cycles d'initiation contrairement à l'unique cycle de réplication initié dans chaque réplicon dans le cycle cellulaire normal. Les régions responsables de l'amplification semblent assez longues, environ 500 pb, comme le prouve l'incapacité de fragments plus courts transloqués en d'autre loci (par transposition à l'aide de l'élément P) de provoquer une amplification.

Le phénotype produit par l'inversion peut être considéré comme un cas particulier d'effet de position. Ce terme décrit l'influence de la localisation sur l'expression d'un gène comme le prouvent les modifications de son activité après translocation. Un exemple bien connu est l'inactivité qui résulte de la translocation d'un gène situé dans l'euchromatine dans le type sauvage vers une nouvelle localisation proche de l'hétérochromatine. Bien que le gène ne soit pas en lui-même modifié, son activité est influencée par les séquences qui l'entourent.

Dans le cas de *ocelliless*, l'inversion provoque la non amplification de certains gènes alors que d'autres sont amplifiés à la place. Les effets de dosage entraînent des changements complexes du phénotype sans altération des gènes en eux-mêmes.

SÉLECTION DE SÉQUENCES GÉNOMIQUES AMPLIFIÉES

Le génome des eucaryotes a la capacité d'assimiler de nouvelles séquences d'origine exogène ou endogène. Qu'elles soient ajoutées aux cellules ou originaires des chromosomes, les séquences d'ADN peuvent être présentes sous forme de multiples copies extrachromosomiques ou chromosomiques qui survivent comme des groupes en tandem. Sous forme extrachromosomique,

le matériel additionnel est transmis de façon irrégulière (il ne ségrège même pas au cours des divisions comme un plasmide authentique). Sous forme chromosomique cependant, le matériel est intégré dans le génome de l'hôte et devient un composant du génotype.

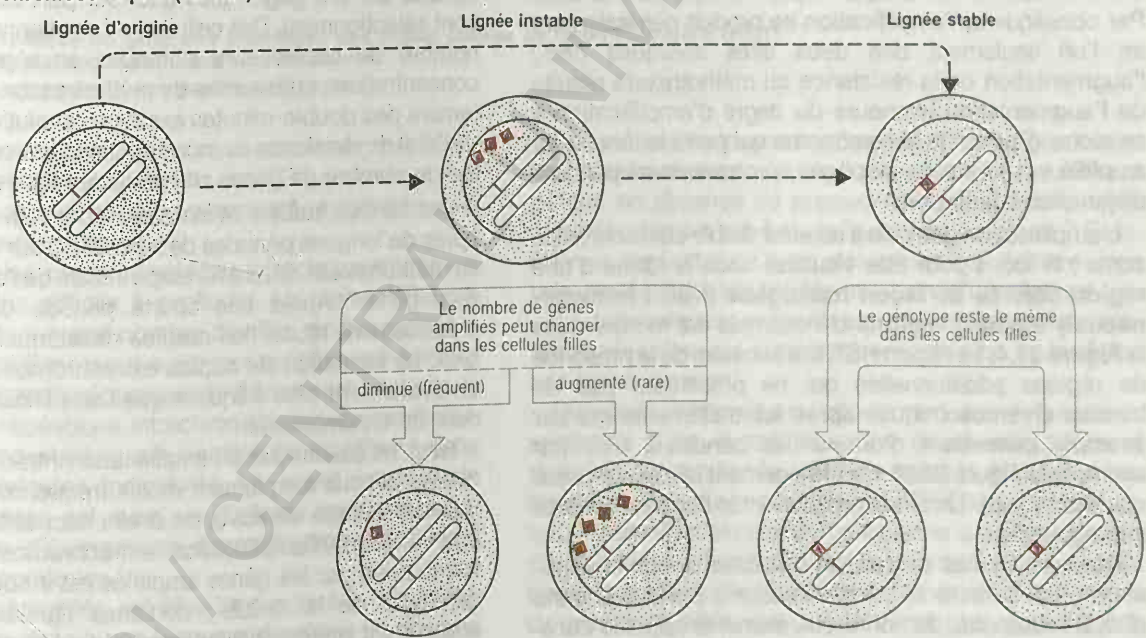
L'amplification de séquences endogènes est provoquée par la sélection de cellules résistantes à certains agents. Le même profil général de flexibilité est observé dans les transfections et dans les amplifications qui très probablement mettent en jeu des mécanismes en partie semblables. Nous ne connaissons rien des activités enzymatiques impliquées dans ces opérations mais nous supposons qu'elles sont engagées dans des activités similaires au cours des étapes normales de la réplication de l'ADN.

L'exemple d'amplification le mieux caractérisé résulte

de l'addition de **méthotrexate** (mtx) à divers types de cellules en culture. Ce réactif bloque le métabolisme des folates. Une résistance à ce composé peut être conférée par des mutations qui changent l'activité de l'enzyme dihydrofolate réductase (DHFR). Une alternative au changement de l'enzyme elle-même est l'augmentation de son activité. L'origine de cette augmentation est l'amplification du nombre de gènes de structure *dhfr*. L'amplification se produit à une fréquence plus grande que la vitesse d'apparition des mutations spontanées, généralement entre 10^{-4} et 10^{-6} . Des événements similaires ont maintenant été observés dans plus de vingt gènes.

Un trait commun à la plupart de ces systèmes est que des cellules hautement résistantes ne sont pas obtenues en une seule étape mais apparaissent par adaptation progressive des cellules à des doses croissantes du

Figure 31.3
Le gène *dhfr* peut être amplifié pour donner des copies additionnelles extrachromosomiques (instables) ou chromosomiques (stables).



Effets des conditions de sélection

Les gènes amplifiés sont nécessaires en présence de méthotrexate, donc les cellules :	meurent s'il y a trop peu de gènes amplifiés	survivent avec des fluctuations quantitatives
Les gènes amplifiés sont un fardeau en absence de sélection, donc les cellules :	envahissent la population	sont éliminées de la population

- Chromosome avec 1 gène DHFR
- Chromosome avec des gènes DHFR amplifiés
- Double minute avec des gènes DHFR amplifiés

réactif toxique. L'amplification génique peut nécessiter plusieurs étapes.

Le nombre de gènes *dhfr* dans une lignée cellulaire résistante au méthotrexate varie de 40 à 400 en fonction de la rigueur de la sélection et de la lignée cellulaire utilisée. Les lignées *mtx^r* se répartissent en deux classes qui se distinguent par la façon dont les cellules poussent en l'absence de méthotrexate, éliminant la pression sélective pour une forte activité DHFR.

Dans les lignées **stables**, les gènes amplifiés sont maintenus. Dans les lignées **instables**, les gènes amplifiés sont au moins en partie perdus lorsque la pression de sélection est levée. L'origine de cette différence réside dans la nature des gènes *dhfr* amplifiés dans ces deux lignées. La situation responsable de ce genre d'état est résumée dans la **Figure 31.3**.

Dans une lignée stable *mtx^r*, les gènes amplifiés *dhfr* sont chromosomiques occupant le site habituel *dhfr* sur l'un des chromosomes. Habituellement, l'autre chromosome conserve son unique copie normale au site *dhfr*. Par conséquent, l'amplification se produit généralement en l'un seulement des deux sites alléliques *dhfr*; l'augmentation de la résistance au méthotrexate résulte de l'augmentation ultérieure du degré d'amplification à ce locus. (Parfois, le chromosome qui porte le locus *dhfr* amplifié est lui-même dupliqué, probablement par non disjonction à la mitose.)

L'amplification génique a un effet visible sur le chromosome ; le locus peut être visualisé sous la forme d'une **région colorée de façon homogène (HSR : homogeneously staining region)**. Un exemple est montré dans la **Figure 31.4**. La région HSR tire son nom de la présence de régions additionnelles qui ne présentent pas de bandes chromosomiques après les traitements qui par exemple permettent d'obtenir les bandes G (voir par exemple la Figure 25.9). Ce changement laisse supposer que les régions chromosomiques inter-bandes ont subi une expansion.

Dans les lignées de cellules instables, aucun changement n'est détecté sur le chromosome portant le gène *dhfr*. En revanche, de nombreux éléments appelés **chromosomes double-minute** sont visibles. Un exemple est montré **Figure 31.5**. Dans une lignée cellulaire typique, chaque double-minute porte 2 à 4 gènes *dhfr*. Les double-minutes semblent se répliquer de façon autonome ; mais ils n'ont pas de centromère. En conséquence ils ne s'attachent pas au fuseau mitotique et ségrègent donc de façon erratique. Il sont fréquemment perdus dans les cellules filles. Nonobstant leur nom, les double-minutes sont considérés comme étant extrachromosomiques.

La transmission irrégulière des double-minutes expli-

que l'instabilité de la résistance au méthotrexate dans ces lignées. Les double-minutes vont être progressivement perdues au cours des divisions cellulaires ; en présence de méthotrexate, les cellules ayant un nombre réduit de gènes *dhfr* mourront. Seules les cellules ayant conservé un nombre suffisant de double-minutes apparaîtront dans la population qui survit.

La présence des double-minutes réduit la vitesse de prolifération des cellules. Par conséquent lorsque la pression de sélection est levée, les cellules qui n'ont pas de gènes amplifiés sont avantagées ; elles engendrent des descendants plus rapidement et envahissent rapidement la population. Cela explique pourquoi la résistance au méthotrexate n'est retenue dans les lignées cellulaires que lorsque les cellules sont maintenues en présence de méthotrexate.

A cause de la ségrégation erratique des double-minutes, l'accroissement du nombre de copies peut survenir relativement rapidement car à chaque cycle cellulaire les cellules qui ont gagné plus que leur part de gènes *dhfr* sont sélectionnées. Des cellules qui contiennent un grand nombre de copies sont obtenues en réponse à des concentrations croissantes de méthotrexate. Le comportement des double-minutes explique l'évolution par étape de l'état de résistance au *mtx^r* et les fluctuations incessantes du nombre de gènes *dhfr* dans les lignées instables.

Les lignées stables et instables sont toutes produites après de longues périodes de sélection pour la résistance au méthotrexate. Quel est l'étape initiale de l'amplification d'un gène ? Après une courte période de sélection, pratiquement toutes les cellules résistantes sont instables. La formation de copies extrachromosomiques est un événement plus fréquent que l'amplification à l'intérieur du chromosome.

Nous ne savons pas si l'amplification intrachromosomique se produit simplement moins fréquemment, constituant une étape de *novo*, ou si elle nécessite une étape intermédiaire d'amplification extrachromosomique. La forme prise par les gènes amplifiés est influencée par le génotype de la cellule ; certaines lignées cellulaires engendrent préférentiellement des double-minutes alors que d'autres exhibent plutôt la configuration HSR.

Comment les copies extrachromosomiques sont-elles produites ? Nous savons que leur production n'entraîne pas la disparition de la copie chromosomique d'origine du gène. Un modèle est illustré sur la **Figure 31.6**. Des cycles additionnels de réplication sont initiés au voisinage du gène *dhfr*. Les copies supplémentaires pourraient être libérées du chromosome par des événements similaires à une recombinaison. En fonction de la nature de ces événements, une molécule extrachromosomique d'ADN,

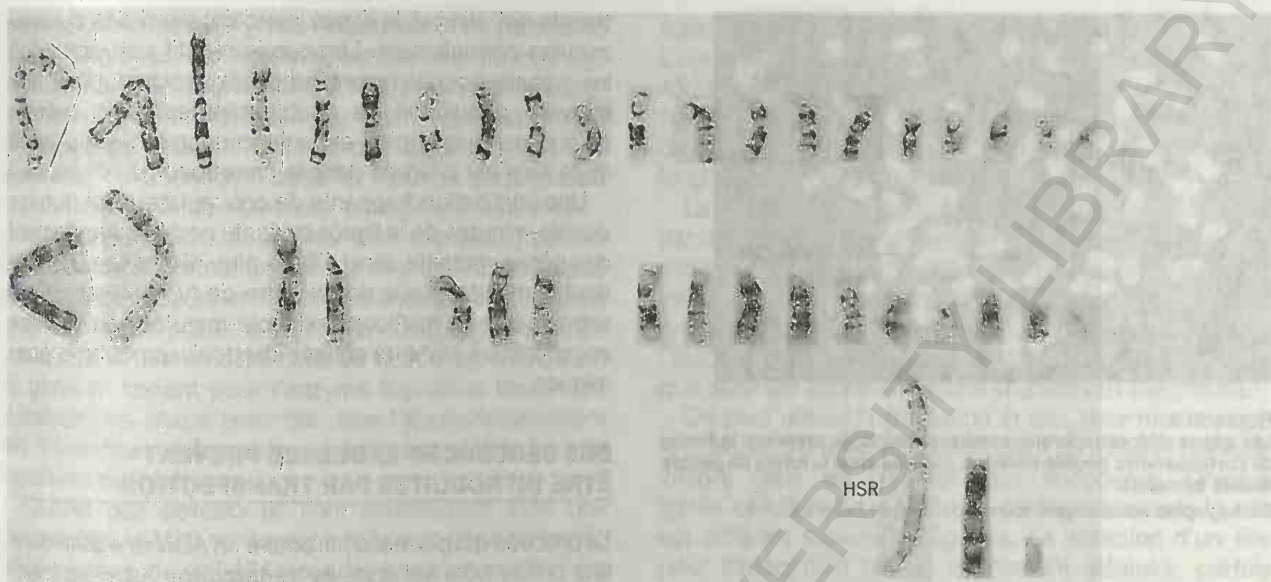


Figure 31.4

Les copies amplifiées du gène *dhfr* produisent une région colorée de façon homogène (HSR : homogeneously staining region) dans le chromosome.

Photographie aimablement fournie par Robert Schimke.

contenant une ou plusieurs copies serait engendrée. Si les double-minutes contiennent des ADN circulaires, des recombinaisons entre eux pourraient engendrer des molécules multimériques.

L'origine des copies amplifiées intrachromosomiques peut être plus difficile à expliquer. La région HSR est constituée d'un grand nombre d'unités répétées en tandem. Un problème pour supposer qu'elles proviennent de l'intégration de séquences double-minutes est que le nombre de répétitions intrachromosomiques est beaucoup plus important que le nombre de répétitions dans les double-minutes individuels. Le grand nombre de gènes intégrés est aussi un obstacle pour construire des modèles d'événements d'amplification strictement intrachromosomiques.

La région amplifiée est plus longue que le gène *dhfr* lui-même. Le gène a une longueur d'environ 31 kb mais la longueur moyenne des unités de répétition est de 500 à 1000 kb dans le HSR chromosomique. La quantité d'ADN contenue dans un double-minute est comprise entre 100 et 1000 kb.

L'étendue de la région amplifiée est variable selon le type cellulaire. L'on peut supposer qu'elle comporte toujours l'origine du réplicon qui contient les gènes amplifiés, mais qu'elle a des extrémités variables.

* N. d. T. : En réalité l'amplification peut se produire au niveau d'un gène *dhfr* cloné et réintroduit par transfection de la cellule ; dans ce cas pourtant le site d'intégration du gène exogène se fait au hasard.

Les mêmes événements d'amplification se produisent probablement sans méthotrexate, mais les produits sont perdus en absence de pression sélective. Pourquoi un tel phénomène n'est-il vu qu'avec peu de gènes ? Peut-être ne survient-il qu'avec des gènes qui résident dans une localisation appropriée au sein d'un réplicon, ou quand des séquences répétées appropriées sont présentes en des sites qui autorisent une excision de la région par recombinaison*.

Quelques informations concernant les événements impliqués dans le maintien des double-minutes sont données par une lignée cellulaire instable dont les gènes amplifiés codent pour une enzyme DHFR mutante. L'enzyme mutante n'est pas présente dans la cellule diploïde d'origine (donc la mutation a dû se produire au cours du processus d'amplification). Malgré la variation du nombre de gènes amplifiés, ces cellules n'ont plus *que* l'enzyme mutée. Donc les gènes chromosomiques de type sauvage ne peuvent engendrer en permanence un grand nombre de nouveaux double-minutes puisque si tel était le cas les copies amplifiées produiraient l'enzyme normale.

Une fois qu'un gène se trouve amplifié sous forme de copies extrachromosomiques, *des changements dans l'état de la cellule sont induits par ces gènes et non par les copies chromosomiques d'origine*. Quand le métho-

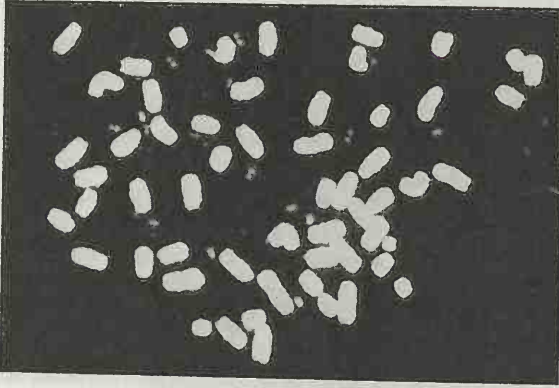
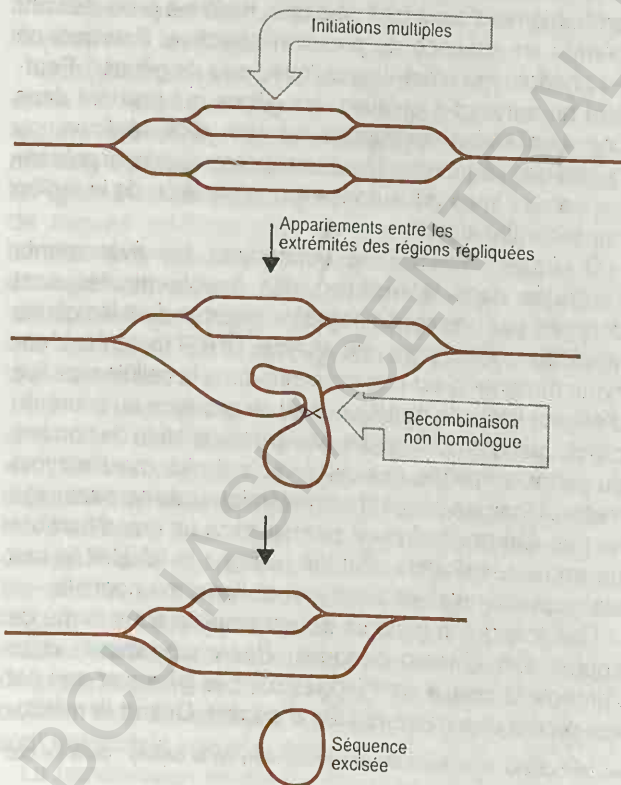


Figure 31.5
Les gènes *dhfr* extrachromosomiques amplifiés prennent la forme de chromosomes double-minutes, visibles sous la forme de petites taches blanches.

Photographie aimablement fournie par Robert Schimke.

Figure 31.6
Les copies extrachromosomiques d'ADN pourraient être engendrées par la réinitiation de la réplication suivie d'une recombinaison non homologue entre les répliques.



trexate est éliminé la lignée cellulaire perd ses double-minutes normalement. Lorsque le réactif est réintroduit les gènes *normaux* sont amplifiés pour donner une nouvelle population de double-minutes. Ceci montre qu'aucune des copies extrachromosomiques du gène muté ne s'est intégrée dans le chromosome.

Une implication frappante de ces résultats est que les double-minutes de la lignée mutante portent *uniquement* des gènes mutants, ainsi s'il y a plus d'un gène *dhfr* par double-minutes, tous doivent être de type mutant. Ceci suggère que les multicopies double-minutes peuvent être engendrées à partir de gènes individuels extrachromosomiques.

DES SÉQUENCES EXOGÈNES PEUVENT ÊTRE INTRODUITES PAR TRANSFECTION

Le procédé qui permet d'introduire un ADN exogène dans une cellule hôte est appelée **transfection**. Les expériences de transfection ont débuté par l'addition de préparations de chromosomes en métaphase à des suspensions cellulaires. Les chromosomes rentrent de façon assez inefficace dans les cellules et conduisent à des variants instables avec une fréquence faible. Des chromosomes intacts survivent rarement au processus ; la cellule hôte conserve habituellement un fragment du chromosome introduit (qui est instable car il n'a pas de centromère). De rares cas de lignées stables peuvent avoir résulté d'une intégration du matériel donneur dans un chromosome hôte.

Des résultats similaires sont obtenus quand de l'ADN purifié est ajouté à une préparation cellulaire réceptrice. Cependant avec de l'ADN purifié il est possible d'ajouter des séquences particulières au lieu de compter sur la fragmentation au hasard des chromosomes. La transfection d'ADN donne des lignées cellulaires stables et instables, les premières étant relativement plus fréquentes. (Ces expériences sont tout à fait analogues à la transformation bactérienne, comme cela est indiqué sur la Figure 2.3, mais sont décrites en tant que transfection du fait de l'utilisation historique du terme transformation pour décrire les changements qui induisent la croissance illimitée des cellules eucaryotes.)

Les faibles rendement de transfection rendent nécessaire l'utilisation des marqueurs dont la présence permet de sélectionner la cellule réceptrice. Il est à noter que la séquence transférée est exprimée. La plupart des expériences de transfection sont réalisées avec des marqueurs représentant des fonctions enzymatiques facilement détectables, mais en principe, tout marqueur qui peut conférer un avantage sélectif peut être utilisé. Ceci

permet l'isolement de gènes responsables de caractères morphologiques. Par exemple, comme nous l'avons déjà mentionné, des cellules transfectées peuvent être sélectionnées pour l'acquisition du phénotype transformé (tumorigène). Ensuite nous pouvons identifier l'ADN responsable du phénotype observé. Ce type de protocole a permis d'isoler de nombreux *oncogènes* cellulaires.

La cotransfection avec plus d'un marqueur, permet d'obtenir des informations sur les événements impliqués dans la transfection, et a étendu le champ des questions que l'on peut poser à l'aide de cette technique. Un marqueur souvent utilisé dans ce type d'expérience est le gène *tk* codant pour l'enzyme thymidine kinase qui catalyse une étape essentielle dans l'approvisionnement en thymidine triphosphate, précurseur utilisé pour la synthèse de l'ADN.

Quand des cellules *tk* sont transfectées avec une préparation d'ADN contenant à la fois un gène *tk*⁺ purifié et le génome de ϕ X 174, tous les transformants *tk*⁺ ont les deux séquences exogènes. C'est une observation utile, parce qu'elle autorise l'introduction de marqueurs qui ne peuvent être sélectionnés, par simple cotransfection avec un marqueur qui peut l'être.

L'arrangement des séquences *tk* et ϕ X 174 est propre à chaque lignée transfectée mais reste identique au cours de la propagation d'une lignée. Souvent de multiples copies de l'ADN donneur sont présentes, le nombre variant suivant les lignées individuelles. Les révertants perdent simultanément les deux séquences ϕ X 174 et *tk*. Sous pression sélective, l'amplification de séquences transfectées induit l'accroissement du nombre de copies de toutes les séquences introduites. Donc les deux types de séquences introduites restent physiquement liées au cours de la transfection et par conséquent subissent ultérieurement le même destin.

Pour réaliser une expérience de transfection, la masse d'ADN ajoutée aux cellules réceptrices est augmentée par addition d'un « ADN carrier » (ADN de transport) qui consiste en une préparation d'un quelconque autre ADN (souvent de sperme de saumon). Les cellules transfectées (souvent de sperme de saumon). Les cellules transfectées se révèlent contenir des séquences « d'ADN carrier » entourant la séquence sélectionnée. Par conséquent la transfection semble mettre en jeu une grande unité constituée de toutes les séquences contenues dans la préparation donneuse, liées en tandem.

Puisque des révertants pour le marqueur de sélection perdent tout ce matériel, il semble probable que la cellule

transfectée acquière uniquement une grande unité. L'unité peut être formée par liaison en concatémère de séquences donneuses dans une réaction qui est rapide comparée aux autres événements qui se produisent lors de la transfection. Cet ensemble transfecté doit avoir une longueur de l'ordre de 1000 kb.

La taille de l'unité transfectée ne permet pas de savoir par des expériences de transfert (blot) si elle est physiquement liée à l'ADN chromosomique de l'hôte (les extrémités du fragment concerné sont présentes en trop petite proportion). Il semble plausible que la première étape soit l'établissement d'une unité instable extrachromosomique, suivi de l'acquisition d'une stabilité via l'intégration.

On peut utiliser l'hybridation *in situ*, pour montrer que les cellules transfectées contiennent du matériel donneur intégré dans les chromosomes endogènes. Chaque lignée cellulaire a un seul site d'intégration mais le site est différent suivant les lignées. La sélection d'un site pour l'intégration est un événement aléatoire, parfois associé à un réarrangement chromosomique massif*.

Quelques résultats intéressants obtenus par cotransfection de deux marqueurs de sélection sont résumés sur la Figure 31.7. Une lignée cellulaire réceptrice *aprt* *tk*⁻ fut transfectée avec de l'ADN portant les gènes *aprt* et *tk*. Le gène *aprt* était l'allèle actif *aprt*⁺ qui code pour l'enzyme adénosine phosphoribosyltransférase. Le gène *tk* était l'allèle inactif contenant la région codante pour la thymidine kinase mais pas de promoteur. Les cellules transfectées furent ensuite sélectionnées pour la fonction *aprt*⁺, mais les deux séquences donneuses furent retrouvées.

Ensuite, les cellules *aprt*⁺ *tk*⁻ furent sélectionnées pour leur capacité à produire la fonction *tk*⁺. Cette acquisition dépend de deux événements. La séquence *tk* doit être transcrite à partir d'un site se substituant à son propre promoteur et comme cela n'entraîne qu'une expression faible, le nombre de copies doit être amplifié. Les cellules sélectionnées deviennent *aprt*⁺ *tk*⁺ et elles ont environ 40 copies de l'unité transfectée *aprt*-*tk*, intégrée en un site chromosomique unique.

Finalement ces cellules sont sélectionnées pour le phénotype *aprt* *tk*⁺. Elles doivent donc éliminer l'expression des multiples copies du gène *aprt*⁺ sans affecter l'expression du gène *tk*. Le fait intéressant est que toutes les copies du gène *aprt* acquièrent dans ces conditions la même mutation négative. Ce résultat implique l'existence d'un mécanisme qui autorise l'ensemble du groupe

* N. d. T. : Les sites d'intégration peuvent parfois résulter de phénomènes de recombinaisons homologues entre les séquences injectées et les séquences endogènes. Un tel phénomène qui permet « d'insérer un gène à sa place » est la règle chez la levure. Dans les cellules de mammifère il semble survenir dans un cas sur 1 000, soit de 100 à 1000 fois plus fréquemment que ne le voudrait l'insertion totalement au hasard d'un fragment d'ADN dans un génome de 310⁹ pb.

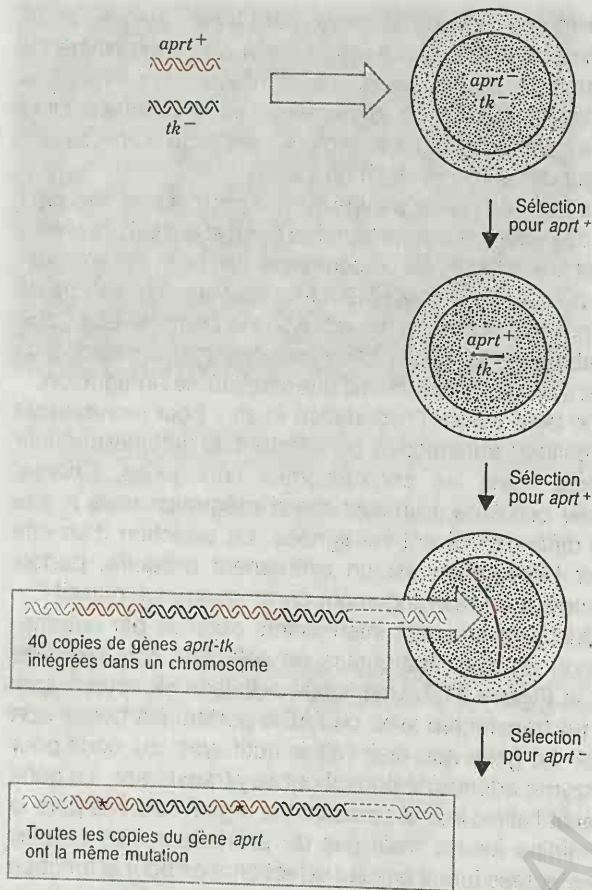


Figure 31.7

Quand des gènes transfectés sont amplifiés, le groupe entier répond de façon coordonnée aux pressions sélectives ultérieures.

de gènes à être régénéré à partir d'un seul membre ou corrigé de façon à refléter la séquence de l'un d'entre eux. Apparemment cela se produit en une seule génération cellulaire.

L'ADN TRANSFECTÉ PEUT PÉNÉTRER DANS LES CELLULES GERMINALES

Un développement passionnant des techniques de transfection est leur application pour introduire des gènes dans des animaux. Un animal qui acquiert une nouvelle information génétique par addition d'un ADN étranger est dit **transgénique**.

* N. d. T. : Actuellement la tendance est de n'injecter que le gène étudié, débarrassé de toute séquence du vecteur plasmidique. Dans un certain nombre de cas, en effet, il semble que la présence d'ADN procaryotique inhibe l'expression d'un gène encaryote à côté duquel il est inséré. La richesse des séquences procaryotiques en doublets CpG, cible possible des méthylations, pourrait expliquer cet effet inhibiteur.

La première question que nous nous posons à propos des animaux transgéniques est de savoir combien de copies du matériel étranger il contient, où sont localisés ces copies et si elles sont présentes dans la lignée germinale et transmises selon un mode mendélien.

Une question importante qui peut être résolue par l'utilisation d'animaux transgéniques concerne l'indépendance des gènes et l'effet des régions dans lesquelles ils résident. Si nous prenons un gène, en incluant les régions bordantes qui contiennent ses éléments connus de régulation, sera-t-il exprimé de la même façon quelle que soit sa localisation dans le génome ? En d'autres termes, les éléments régulateurs fonctionnent-ils indépendamment ou bien l'expression d'un gène est-elle aussi contrôlée par d'autres facteurs tels par exemple la localisation dans un domaine approprié d'un chromosome ?

Une approche qui s'est révélée couronnée de succès avec la souris est résumée sur la Figure 31.8. Des plasmides contenant le gène d'intérêt sont injectés dans l'un des pronuclei d'un œuf fécondé*. L'œuf est réimplanté dans une femelle pseudogestante. Après la naissance, la souris réceptrice peut être examinée pour voir si elle a vraiment acquis l'ADN étranger et si elle l'exprime.

Des séries d'expériences ont été réalisées chez la souris avec un gène de fusion qui a le promoteur de la méthallothionéine lié à la séquence codante pour la thymidine kinase. Le promoteur MT est dérivé d'un gène naturel de souris et peut être testé quant à sa capacité à répondre aux inductions habituelles par les glucocorticoïdes et les métaux lourds. La réponse peut être testée en mesurant l'activité thymidine kinase.

Le résultat usuel de telles expériences est qu'une minorité raisonnable (disons environ 15 %) des souris injectées porte la séquence transfectée. Habituellement, de multiples copies du plasmide sont intégrées en tandem en un site chromosomique unique. Le nombre de copies varie de 1 à plusieurs dizaines. Ils sont transmis à la descendance de la souris injectée comme on l'attend d'un locus mendélien.

Les gènes transfectés s'expriment-ils avec la spécificité correcte au cours du développement ? La règle générale apparaît être aujourd'hui qu'il y a un fac-similé raisonnable des contrôles corrects : les gènes transfectés sont généralement exprimés dans les cellules appropriées et au moment voulu. Il existe cependant des exceptions dans lesquelles un gène transfecté est exprimé d'une façon non physiologique dans certains tissus :

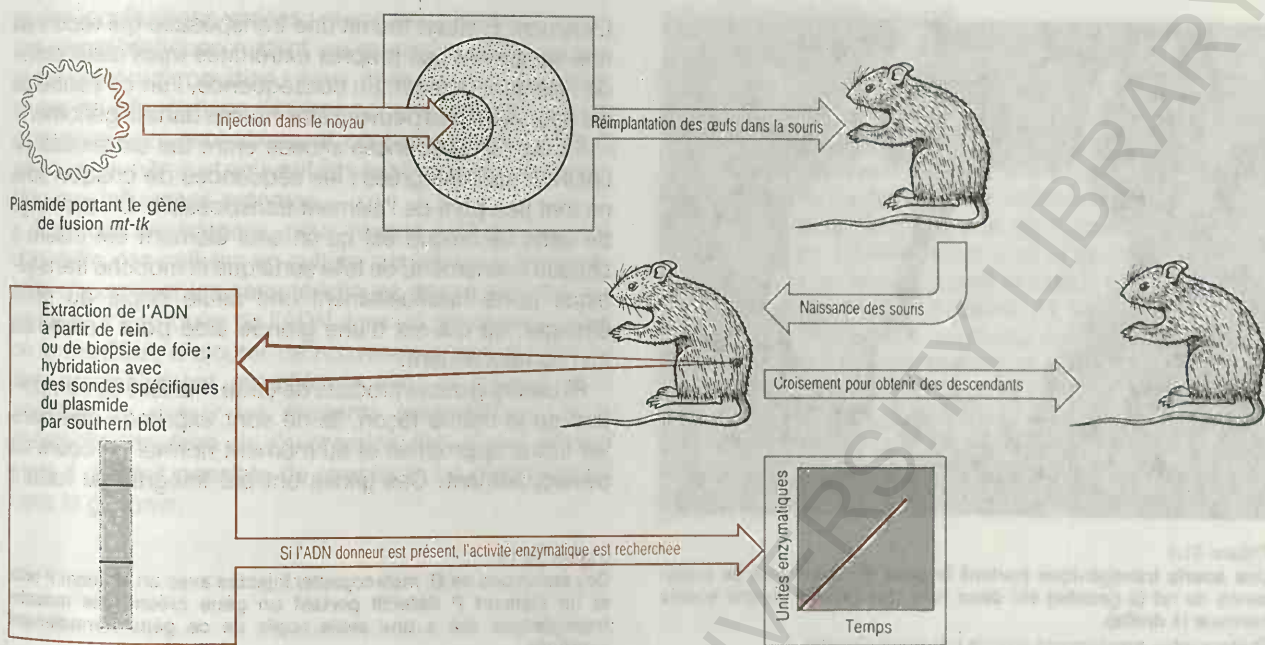


Figure 31.8
Par injection, de l'ADN peut être introduit dans les lignées germinales des animaux.

Dans la descendance de la souris injectée, l'expression du gène donneur est extrêmement variable* ; elle peut être complètement éteinte, quelque peu réduite ou même augmentée. Même chez le parent fondateur, le niveau d'expression du gène n'est pas corrélé au nombre de copies intégrées en tandem. Probablement un nombre réduit de ces gènes est actif. (Il faut aussi noter que si tous les gènes additionnels étaient exprimés il serait surprenant d'obtenir une régulation correcte puisque le grand nombre de promoteurs entraînerait une dilution des molécules régulatrices.)

Quelles peuvent être les causes de la variation de l'expression des gènes ? Une possibilité qui a souvent été discutée pour des gènes transfectés (et qui s'applique aussi aux génomes rétroviraux intégrés) est que le site d'intégration peut être important. Un gène pourrait être exprimé lorsqu'il s'est intégré dans un domaine chromatinien actif mais pas lorsqu'il le serait dans une autre région**. Une autre possibilité est la survenue de modi-

fications épigénétiques ; par exemple des changements du profil de méthylation pourraient être responsables des changements dans l'activité. Alternativement, le gène actif chez le parent pourrait être délété ou amplifié dans la descendance.

Un exemple particulièrement intéressant des effets d'un gène injecté est fourni par la lignée de souris transgéniques dérivée d'œufs injectés avec un gène de fusion composé du promoteur MT lié au gène de structure de l'hormone de croissance de rat. Les niveaux d'hormone de croissance chez certaines des souris transgéniques étaient plusieurs centaines de fois supérieurs à la normale. Les souris devinrent presque deux fois plus grosses que la normale comme le montre la **Figure 31.9**.

L'introduction d'oncogènes peut induire la formation de tumeurs. Des souris transgéniques exprimant les régions précoces de SV40 et les éléments régulateurs, expriment les gènes viraux qui codent pour les antigènes grand T et petit t dans seulement quelques tissus le plus

* **N. d. T. :** Si on exclut les cas où le fondateur transgénique est une mosaïque comportant plusieurs sites d'intégration du transgène (et dans ce cas, ces sites peuvent ségréger dans la descendance), le profil d'expression du transgène est en fait assez stable dans toute la descendance d'un animal transgénique fondateur.

** **N. d. T. :** La création d'animaux transgéniques pour le gène de β -globine, contenant un élément situé très en amont de ce gène, aboutit à l'expression du transgène de façon indépendante du site d'intégration et proportionnelle au nombre de copies. On pense que l'élément amont compris dans le transgène suffit à commander la constitution d'une structure chromatinienne active dans les cellules érythroïdes, et ce de façon dominante par rapport à l'influence du génome environnant.

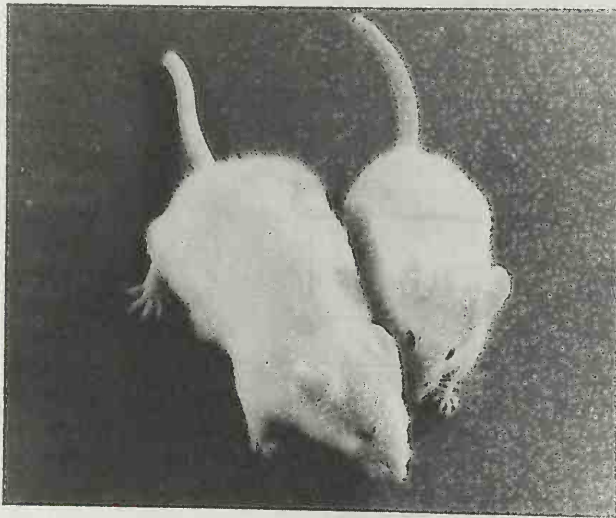


Figure 31.9

Une souris transgénique portant le gène de l'hormone de croissance de rat (à gauche) est deux fois plus grosse qu'une souris normale (à droite).

Photographie aimablement fournie par Ralph Brinster.

souvent le cerveau, le thymus et le rein. Ces souris meurent habituellement avant six mois des conséquences d'une tumeur cérébrale. Parfois les tumeurs ne sont trouvées que dans le thymus et le rein. Dans les lignées cellulaires tumorales, la région SV40 est souvent amplifiée d'un facteur allant jusqu'à 20.

Bien que SV40 soit connu pour provoquer des tumeurs chez le hamster, il n'avait pas été jusqu'alors supposé être oncogénique chez la souris. Dans la lignée transgénique cependant les antigènes T et t de SV40 se comportent comme les produits d'un oncogène intégré. Ces souris peuvent fournir un modèle extrêmement utile pour étudier l'origine de ce type de cancer. D'autres oncogènes peuvent être employés pour engendrer des souris développant d'autres cancers, rendant possible l'obtention de nombreux modèles. Par exemple l'introduction du gène *myc* sous contrôle d'un promoteur actif provoque l'apparition d'adénocarcinomes et autres tumeurs*.

Une méthode plus sophistiquée pour introduire de nouvelles séquences d'ADN a été développée avec *D. melanogaster* en tirant avantage de l'élément P. Le protocole est illustré sur la Figure 31.10. Un élément P déféctif portant le gène d'intérêt est injecté avec un élément P intact dans des embryons préblastodermiques.

* N. d. T. : De nombreuses expériences ont été réalisées dans lesquelles l'expression d'oncogènes est contrôlée par des promoteurs spécifiques de différenciation tissulaire. Habituellement ces transgènes hybrides entraînent des tumeurs du tissu dans lesquels s'expriment normalement les séquences de contrôle utilisées.

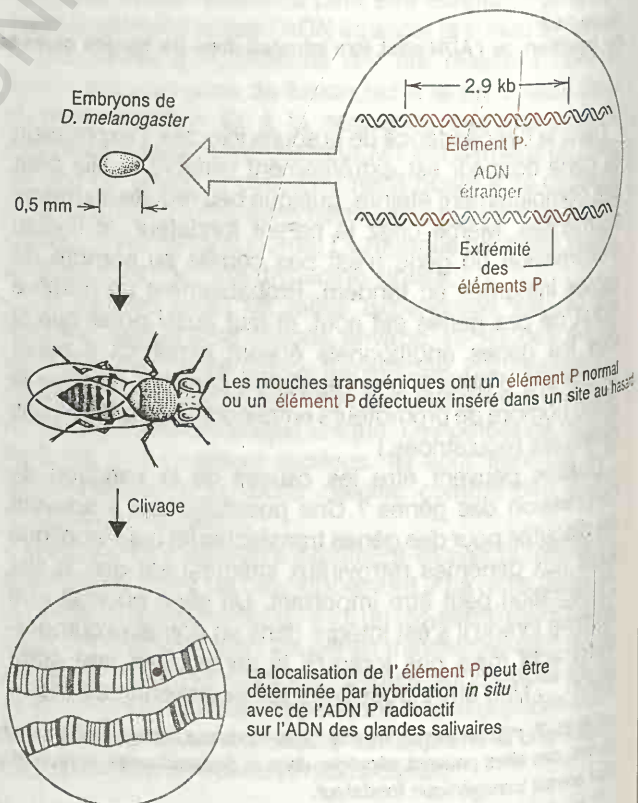
L'élément P intact fournit une transposase qui reconnaît non seulement ses propres extrémités mais aussi celles de l'élément déféctif. En conséquence, l'un ou l'autre ou les deux éléments peuvent être insérés dans le génome.

Seules les séquences situées entre les extrémités de l'ADN P sont intégrées ; les séquences de chaque côté ne font pas parti de l'élément transposable. Un avantage de cette technique est qu'un seul élément est inséré à chaque événement, de telle sorte que la mouche transgénique porte habituellement une seule copie du gène étranger, ce qui est d'une grande aide pour l'étude de son comportement.

Plusieurs gènes introduits de cette manière se comportent de la même façon. Ils ne sont exprimés que dans les tissus appropriés et au moment normal au cours du développement. Ces gènes ont été intégrés au hasard

Figure 31.10

Des embryons de *D. melanogaster* injectés avec un élément P actif et un élément P déféctif portant un gène créent une mouche transgénique qui a une seule copie de ce gène normalement exprimée.



en des localisations variées ; et encore plus passionnant, ces caractéristiques restent vraies pour un gène intégré dans l'hétérochromatine. Ainsi chez *D. melanogaster*, toutes les informations requises pour réguler l'expression d'un gène sont probablement contenues dans le locus génique lui-même et peuvent être relativement imperméables aux influences externes.

Avec ces expériences, nous voyons qu'il est possible d'étendre des cellules en culture aux animaux, la possibilité d'examiner les caractéristiques des régulations. La capacité d'introduire de l'ADN dans un génotype permet de le modifier, d'ajouter de nouveaux gènes qui ont pu être préalablement modifiés *in vitro*. Il devient ainsi possible de cerner les éléments responsables de la spécificité tissulaire de l'expression des gènes. Finalement il devient possible de remplacer un gène défectif dans le génome.

POUR EN SAVOIR PLUS

Le génome de *D. melanogaster* fut passé en revue par **Spradling & Rubin** (*Ann. Rev. Genet.* **15**, 219-264, 1981) ; les effets de la mutation *ocelliless* furent découverts par **Spradling, Spradling & Mahowald** (*Cell* **27**, 193-202 et 203-210, 1981). Les événements impliqués dans l'amplification de *dhfr* ont été passés en revue par **Schimke et al.** (*Cold Spring Harbor Symp. Quant. Biol.* **45**, 785-797, 1981) et (*Cell* **37**, 367-379, 1984) ; d'autres systèmes et mécanismes possibles ont été revus par **Stark & Wahl** (*Ann. Rev. Biochem.* **53**, 447-491, 1984). Les utilisations et les perspectives offertes par les transfections ont été passées en revue par **Pellicer et al.** (*Science* **209**, 1414-1422, 1980). Une série d'expériences d'injection de gènes dans les souris a été passée en revue par **Palmiter & Brinster** (*Cell* **41**, 343-345, 1985). La transfection à l'aide des éléments P fut développée par **Spradling & Rubin** (*Science* **218**, 341-353, 1982).

PARTIE 10

LES GÈNES AU JOURS DU DÉVELOPPEMENT

Édition Springer 1985

BCU IASI / CENTRAL UNIVERSITY LIBRARY



CHAPITRE 32

RÉARRANGEMENTS ET ÉVOLUTION

PARTIE 10

LES GÈNES AU COURS DU DÉVELOPPEMENT

En considérant la structure des fibres chromosomiques comme un scénario codé, on suppose que l'intelligence la plus vive pourrait dire à partir de leur structure, si l'œuf se transformera, dans des conditions favorables, en un coq noir ou en une poule tachetée, en une mouche ou en un plant de maïs, un scarabée, une souris ou une femme... Mais le terme de « scénario codé » est bien sûr trop étroit. Les structures chromosomiques contribuent en même temps à déterminer le développement qu'elles laissent prévoir. Elles représentent en même temps le pouvoir législatif et le pouvoir exécutif ou, en d'autres termes, elles sont le plan de l'architecte et le constructeur.

Erwin Schrödinger, 1945

BCU IASI / CENTRAL UNIVERSITY LIBRARY

LES GÈNES DU DÉVELOPPEMENT

PARTIE 10

Le développement est un processus complexe qui implique l'interaction de nombreux gènes. Ces gènes sont organisés en réseaux de régulation, où les produits de gènes agissent comme facteurs de transcription pour réguler l'expression d'autres gènes. Ce processus est contrôlé par des signaux externes et internes, permettant à l'organisme de développer une forme spécifique à partir d'une cellule unique.

CHAPITRE 32

RÉARRANGEMENTS ET CONSTITUTION DE LA DIVERSITÉ IMMUNITAIRE

On admet comme axiome que la constitution génétique d'un zygote, résultant de la combinaison d'un spermatozoïde et d'un œuf, est héritée par toutes les cellules somatiques de l'organisme. On compte davantage sur un contrôle différentiel de l'expression génique que sur des changements dans le contenu en ADN, pour expliquer les différents phénotypes des cellules somatiques spécifiques.

Toutefois, il existe des situations exceptionnelles dans lesquelles la réorganisation de séquences spécifiques de l'ADN est utilisée pour contrôler l'expression génique ou pour créer de nouveaux gènes. Les gènes des immunoglobulines constituent un cas remarquable et fournissent des changements dans le contenu du génome et de constructions de gènes actifs par recombinaison dans les cellules somatiques appropriées. D'autres cas sont représentés par la substitution d'une séquence par une autre pour modifier le type sexuel de la levure ou pour créer la variété dans l'expression des antigènes de surface grâce auxquels les trypanosomes perpétuent leur cycle infectieux (voir Chapitre 33). La gamme des changements dans le génome eucaryote va donc des transpositions imprévisibles et rares aux reconstructions spécifiques d'un tissu qui se produisent à chaque génération.

La **réponse immunitaire** des mammifères s'exerce grâce à un répertoire apparemment illimité d'**anticorps**, chacun étant synthétisé en réponse à la stimulation par un **antigène** différent. En général l'antigène est une

protéine (ou un résidu attaché à une protéine) d'origine étrangère qui est entrée dans le flux sanguin de l'animal (par exemple, ce peut être le manteau protéique d'un virus). L'exposition à un antigène initie la production d'un anticorps qui *reconnaît spécifiquement l'antigène*.

Les réactions immunitaires sont sous la responsabilité des cellules blanches du sang — les lymphocytes B et T, et les macrophages. Le nom des lymphocytes dérive du tissu dont ils sont originaires. Les **cellules B** sont produites par la moelle osseuse (en anglais : bone marrow) alors que les **cellules T** le sont par le thymus. Chaque classe de lymphocytes utilise le mécanisme de réarrangement de l'ADN pour produire les protéines qui leur permettent de participer à la réponse immunitaire. L'interaction des lymphocytes B et T est décrite dans la **Figure 32.1**.

La production des anticorps est le premier événement responsable de la reconnaissance d'un antigène. Les anticorps sont synthétisés dans les cellules B et leur permettent (en conjonction avec les cellules T et les macrophages) de retirer l'antigène agresseur. Dès qu'un organisme a synthétisé un anticorps spécifique il garde cette capacité de fabrication pour lutter contre toute re-exposition future. Le fait qu'un animal puisse produire un anticorps approprié chaque fois qu'il est exposé à un nouvel antigène, a été l'un des points mystérieux de ce processus. Comment l'organisme peut-il être préparé à produire des protéines anticorps, chacune étant spécifi-

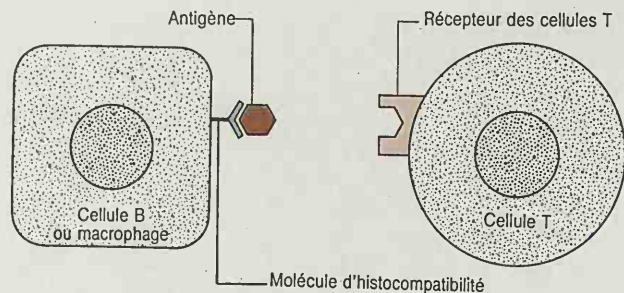


Figure 32.1

Les macrophages ou les cellules B présentant l'antigène dans le contexte des molécules d'histocompatibilité sont reconnus par les cellules T qui utilisent le récepteur cellulaire T pour reconnaître le complexe antigène-molécule d'histocompatibilité.

quement destinée à reconnaître un antigène dont la structure ne peut pas être prévue ?

Les cellules T ont une variété de fonctions, liées aux interactions entre cellules, impliquées dans la réponse immunitaire. La fonction des cellules T implique la production de **récepteurs des cellules T**, un ensemble de protéines qui constituent une contrepartie directe des anticorps produits par les cellules B. Le récepteur permet à une cellule T de reconnaître les cellules B avec lesquelles elle coopérera pour retirer les antigènes du flux sanguin. Il y a moins de variabilité dans les récepteurs des cellules T que dans les anticorps, mais leurs composants sont généralement similaires et leur production emploie les mêmes mécanismes que ceux utilisés pour la synthèse des anticorps.

De nombreuses réactions du système immunitaire sont influencées par une autre série de protéines, les **antigènes d'histocompatibilité**. Chaque individu possède un ensemble caractéristique d'antigènes d'histocompatibilité. Leur nom reflète leur importance dans le rejet des greffes ; une greffe de tissu provenant d'un individu différent est rejetée à cause de la différence existant entre les antigènes d'histocompatibilité du donneur et du receveur, un résultat d'une grande portée médicale. Son ensemble d'antigènes d'histocompatibilité définit une cellule comme « soi » ou « non soi » vis-à-vis d'une autre cellule. Au niveau cellulaire, les antigènes d'histocompatibilité sont impliqués dans les réactions entre lymphocytes.

Chacun de ces trois ensembles de protéines — anticorps, récepteurs des cellules T, antigènes d'histocompatibilité — est diversifié. Si on examine un grand nombre d'individus, on trouvera de nombreuses variantes de chaque protéine. Chaque protéine est codée par une vaste famille de gènes ; et dans le cas des anticorps et du récepteur des cellules T, la diversité de la population

est accrue par des réarrangements de l'ADN qui se produisent dans les lymphocytes impliqués.

Les anticorps et les récepteurs des cellules T ont des structures proches et leurs gènes ont une organisation apparentée. Les antigènes d'histocompatibilité ont également certains caractères en commun avec les anticorps, ainsi que d'autres protéines spécifiques des lymphocytes. En étudiant l'organisation génétique du système immunitaire, on s'intéresse à une série de familles de gènes apparentées, peut-être une « super-famille » qui s'est développée à partir d'un ancêtre commun représentant une réponse immunitaire primitive.

L'importance de la réponse immunitaire pour les mammifères est soulignée par le très grand nombre de gènes contenus dans ces familles ; et elles nous offrent l'opportunité de caractériser les importants groupements de gènes. Éventuellement on devrait pouvoir analyser la réponse immunitaire entière en termes de propriétés des produits des gènes et expliquer l'évolution du système.

L'ORGANISATION DES IMMUNOGLOBULINES

Pour des raisons pratiques, on estime habituellement qu'un mammifère peut produire plus de 10^6 anticorps différents. Chaque anticorps est une **immunoglobuline** tétramérique constituée de deux **chaînes légères (L)** identiques et de deux **chaînes lourdes (H)** (H : heavy en anglais) identiques. Si n'importe quelle chaîne légère peut s'associer avec n'importe quelle chaîne lourde, il faut plus de 1 000 chaînes légères différentes et plus de 1 000 chaînes lourdes différentes pour produire plus de 1 000 000 d'anticorps potentiels.

La structure d'une immunoglobuline tétramérique est illustrée dans la **Figure 32.2**. Chaque chaîne protéique est constituée de deux régions principales : la **région variable (V)** N-terminale ; et la **région constante (C)** C-terminale. À l'origine elles furent définies en comparant les séquences d'acides aminés de différentes chaînes d'immunoglobulines. Comme leurs noms le suggèrent, les régions variables montrent des variations considérables dans leur séquence d'une protéine à l'autre, alors que les régions constantes montrent une importante homologie. Les régions correspondantes des chaînes légères et lourdes s'associent pour former des domaines distincts dans l'immunoglobuline.

Le domaine variable est formé par association entre les régions variables de la chaîne légère et celles de la chaîne lourde. Le domaine V est responsable de la reconnaissance de l'antigène. La production des domai-

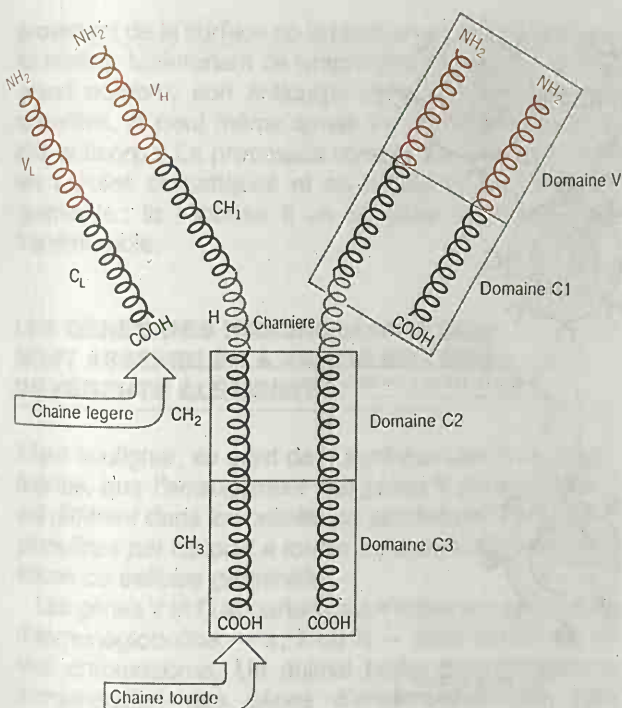


Figure 32.2

Les chaînes lourdes et légères s'associent pour former une immunoglobuline constituée de plusieurs domaines différents.

Chaque chaîne est constituée d'un domaine variable et d'un (chaîne L) ou plusieurs (chaîne H (H = heavy)) domaines constants. Dans la région variable, certaines régions sont plus variables que d'autres. Les parties les plus variables sont appelées **régions hypervariables (HV)**. Les parties relativement moins variables sont les **séquences de base (FR = « framework »)**. Une région variable peut être écrite comme une alternance de ces types de sous-régions : FR1-HV1-FR2-HV2-FR3-HV3-FR4. Les segments hypervariables sont plus courts que les séquences de base (« framework »).

Les régions V avec des spécificités appropriées engendrent la capacité de répondre à divers antigènes. Le nombre total de régions variables pour chacune des chaînes protéiques légère ou lourde se compte en centaines. Ainsi la protéine montre une variabilité maximum dans la région dont la cible diffère à chaque réaction individuelle anticorps-antigène.

L'association des régions constantes engendre plusieurs domaines particuliers dans la molécule. Le premier domaine résulte de l'association de la région constante de la chaîne légère avec la partie CH1 de la région constante de la chaîne lourde. Le reste de la région constante dépend de la seule chaîne lourde. Dans l'exemple montré dans la Figure 32.2, une courte région charnière connecte la première moitié de la molécule aux deux autres domaines constants, chacun formé par

association des régions correspondantes (CH2 et CH3) des chaînes lourdes.

Les domaines constants supportent les fonctions effectrices responsables de l'exécution de la réponse immunitaire. Ainsi les régions de la molécule ayant des fonctions fixes sont conservées. En fait, différentes classes d'immunoglobulines ont des fonctions effectrices apparentées mais pas nécessairement identiques ; le type de la région constante de la chaîne lourde détermine le type de la fonction effectrice.

Les chaînes légères et lourdes sont classées selon leurs régions constantes. Un lymphocyte B donné produit un seul type de tétramères d'immunoglobuline, constitué d'une chaîne légère spécifique combinée à une chaîne lourde spécifique. La chaîne légère peut être l'un des deux types appelés **kappa** (κ) et **lambda** (λ). Les régions constantes **lourdes (H)** constituent un seul groupe. Chacune de ces trois classes (κ , λ et H) possède 1 à 10 membres différents.

En comparant les caractéristiques des régions variables et des régions constantes, on met en évidence un dilemme fondamental dans la structure des gènes des immunoglobulines. Comment le génome code-t-il pour un ensemble de protéines dans lequel toute chaîne polypeptidique doit avoir une des 10 régions constantes possibles, mais peut avoir une des 1 000 régions variables possibles ? Il en résulte que le nombre de séquences codant pour chaque type de région reflète sa variabilité. Il existe de nombreux gènes codant pour les régions V, et seulement quelques gènes codant pour les régions C.

Dans ce contexte, « **gène** » signifie une séquence d'ADN codant pour une partie distincte du polypeptide immunoglobuline final (chaîne lourde ou légère). Ainsi les gènes V codent pour les régions variables et les gènes C codent pour les régions constantes, bien qu'aucun type de gène ne soit exprimé comme unité indépendante. Pour construire une unité qui puisse être exprimée sous la forme d'une authentique chaîne légère ou lourde, un gène V doit être lié physiquement à un gène C. Dans ce système, deux « gènes » codent pour un polypeptide.

Les séquences codant pour les chaînes κ , λ et H sont toutes assemblées de la même manière : n'importe lequel des nombreux gènes V peut être lié à n'importe lequel des quelques gènes C de la même famille. Le processus se produit par **recombinaison somatique** des séquences d'ADN appropriées dans le lymphocyte B dans lequel l'anticorps est exprimé. Dans la plupart des cas, le nombre de gènes V est très élevé, supérieur à 100, organisés en un vaste ensemble de gènes. La variété dans le choix des gènes V est responsable, en grande partie, de la diversité des immunoglobulines. Cependant,

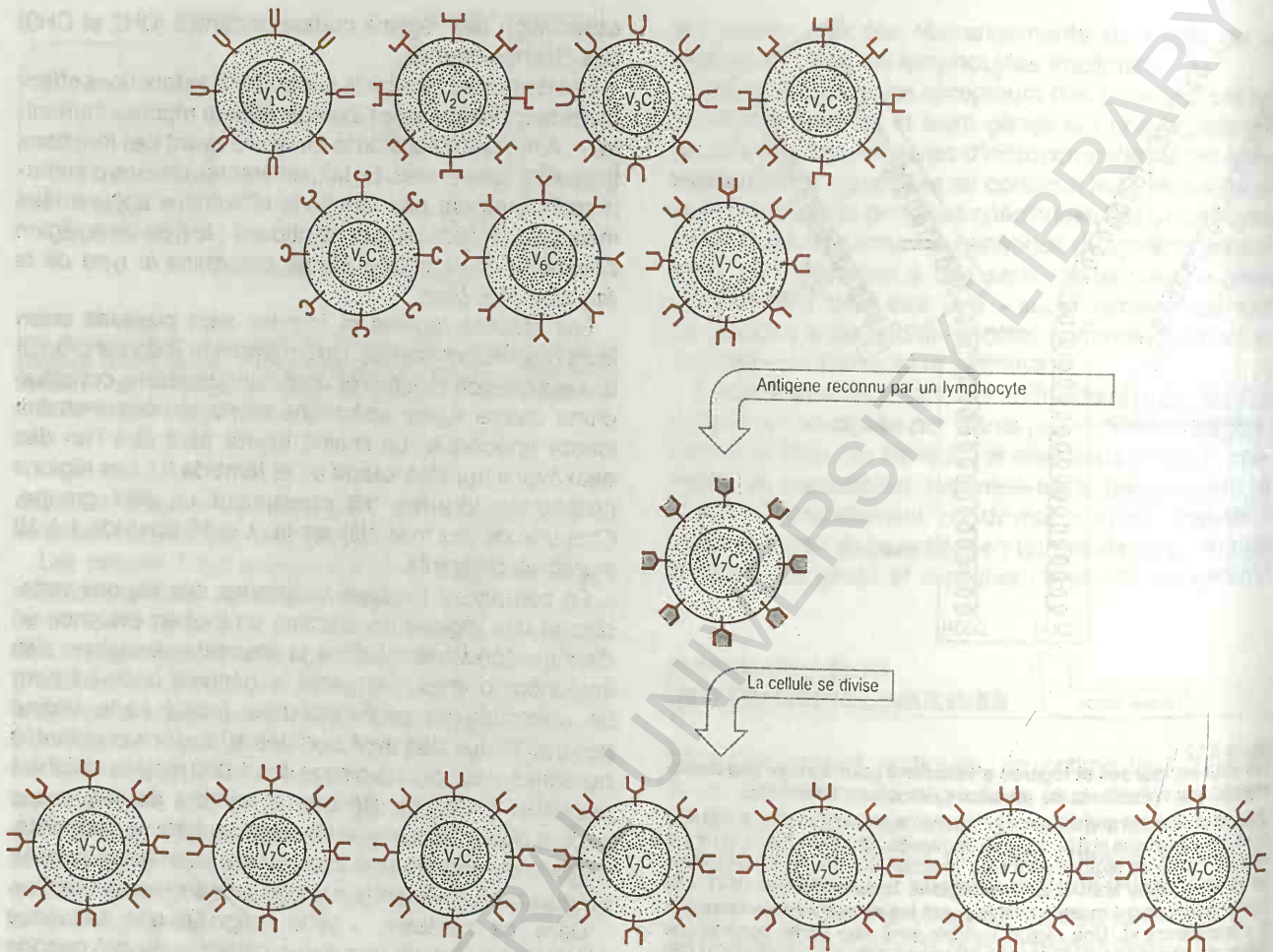


Figure 32.3

La réaction avec un antigène conduit à l'expansion clonale du lymphocyte possédant l'anticorps correspondant.

toute la diversité n'est pas codée dans le génome ; une partie est créée par les changements qui se produisent au cours du processus de construction d'un gène fonctionnel.

La construction d'un gène d'immunoglobuline fonctionnel pourrait ressembler à un processus lamarckien, en ce qu'il représente un changement dans le génome répondant à un caractère particulier du phénotype (l'antigène). A la naissance, l'organisme ne possède pas le gène fonctionnel pour produire un anticorps spécifique. La construction ultérieure du gène permet la synthèse de l'anticorps qui sera ainsi disponible pour réagir avec l'antigène. La rétroaction de l'antigène sur la production de l'anticorps est un phénomène surprenant, mais il peut

être expliqué sans recourir à l'hérésie lamarckienne, grâce à la **théorie de la sélection clonale** résumée dans la Figure 32.3.

La recombinaison des gènes V et C qui donnera des loci fonctionnels se produit dans une population de lymphocytes B immatures. *Chaque cellule produit un seul anticorps* ; elle assume un seul réarrangement productif des gènes des chaînes légères et un seul des gènes des chaînes lourdes. L'anticorps est différent dans chaque cellule, parce que des gènes V et des gènes C différents ont été joints dans chaque reconstruction.

Après contact avec l'antigène, une cellule dont l'anticorps est capable de se lier à l'antigène sera stimulée et se divisera, vraisemblablement à cause d'une rétroaction

provenant de la surface où la réaction anticorps-antigène se réalise. Maintenant ce lymphocyte est produit en très grand nombre, son anticorps est sécrété en grandes quantités, et peut même arriver à dominer la population des anticorps. Le processus complet se manifeste dans les cellules somatiques et ne concerne pas la lignée germinale ; la réponse à un antigène n'est donc pas transmissible.

LES GÈNES DES IMMUNOGLOBULINES SONT ASSEMBLÉS À PARTIR DE LEURS DIFFÉRENTS ÉLÉMENTS

Il faut souligner, au sujet de la synthèse des immunoglobulines, que *l'arrangement des gènes V et des gènes C est différent dans les cellules qui produisent les immunoglobulines par rapport à toutes les autres cellules somatiques ou cellules germinales.*

Les gènes V et C appartenant à n'importe quelle famille d'immunoglobuline — α , λ ou H — sont portés par un seul chromosome. Un animal hérite d'un diagramme d'organisation des gènes d'immunoglobulines dans lequel les gènes V et les gènes C de chaque famille sont très éloignés les uns des autres (on ne sait pas encore exactement à quelle distance). Celui-ci est appelé le **diagramme germinal**, et il est également rencontré dans les cellules somatiques de toute origine autre que le système immunitaire.

Mais dans une cellule exprimant un anticorps, chacune de ses chaînes — un type léger et un type lourd — est codée par un seul gène intact. Le gène actif est engendré grâce à une recombinaison qui apparie un gène V avec

un gène C. Le gène reconstruit inclut les introns, et ses exons correspondent précisément aux domaines fonctionnels de la protéine.

Dans la région variable, le premier exon code pour la séquence signal (impliquée dans l'attachement à la membrane), et le second exon code pour la principale partie de la région variable elle-même (moins de 100 codons).

La structure de la région constante dépend du type de chaîne. Pour les chaînes légères, la région constante est codée par un seul exon (le troisième du gène intact). Pour les chaînes lourdes, la région constante est codée par plusieurs exons ; pour la chaîne protéique montrée dans la Figure 32.2, les différents exons codent pour les régions CH1, charnière, CH2 et CH3. Chaque exon CH mesure environ 100 codons ; la charnière est plus courte. Les introns sont habituellement relativement petits (environ 300 pb).

L'expression des gènes des immunoglobulines procède de la façon habituelle par transcription d'un ARN nucléaire qui est épissé pour donner un ARNm ininterrompu (qui, en plus des régions codantes, contient des régions de tête (leader) et de queue (trailer) non traduites, codées comme parties intégrantes des codons terminaux).

Le nombre de gènes V et de gènes C dans une famille varie avec l'espèce. Le **Tableau 32.1** décrit les constituants de chaque famille d'immunoglobuline chez l'homme et la souris.

Une chaîne légère lambda est assemblée à partir de deux éléments, comme illustré dans la **Figure 32.4**. Le gène V est constitué d'un exon leader (L) séparé par un seul intron du segment variable (V). Le gène C est constitué du segment J séparé par un seul intron de l'exon constant (C).

Tableau 32.1
Chaque famille d'immunoglobuline est constituée d'un groupe de gènes V reliés à leur(s) gène(s) C.

Famille	Localisation sur le chromosome		Nombre de gènes V		Nombre de gènes C	
	Homme	Souris	Homme	Souris	Homme	Souris
Lambda	22	16	< 300	2	> 6	3
Kappa	2	6	< 300	~ 300	1	1
Lourde	14	12	~ 300	> 100	9	8

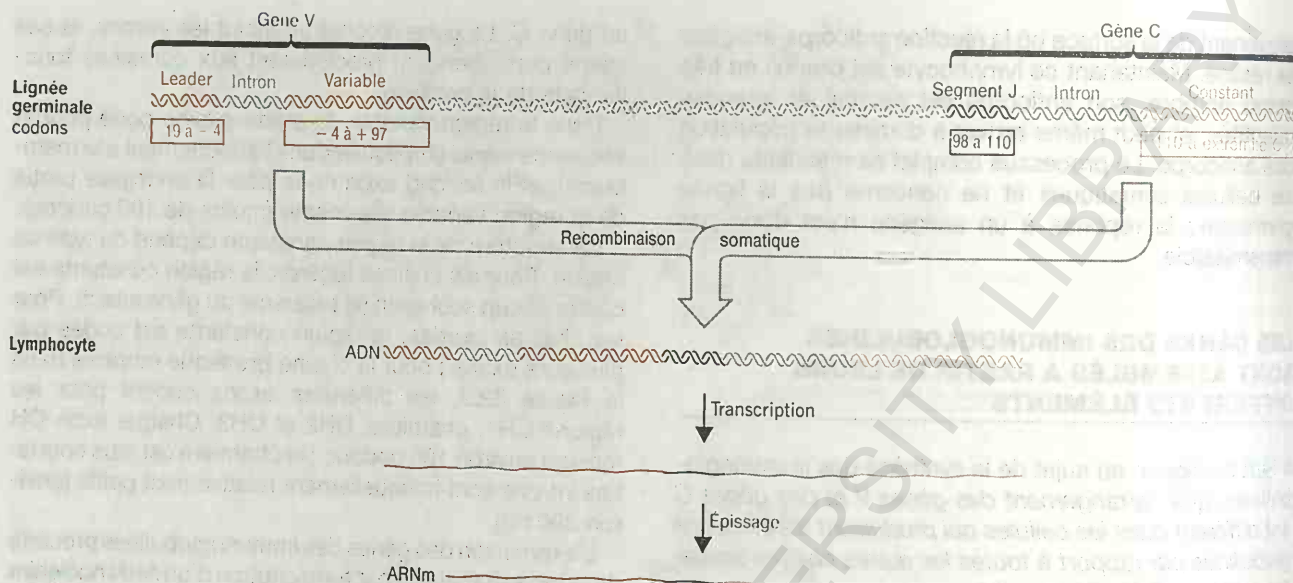


Figure 32.4

Le gène C lambda est précédé d'un segment J, de telle sorte que la recombinaison V-J produit un gène fonctionnel pour une chaîne légère lambda.

Le nom du **segment J** est une abréviation de jonction, puisqu'il identifie la région à laquelle le segment V vient se connecter. Le segment J est court et en fait code pour les tous derniers (13) acides aminés de la région variable, comme on peut le voir dans les séquences d'acides aminés. Dans le gène intact, le segment V-J constitue alors un exon unique codant pour toute la région variable.

Un gène V_λ peut être attaché à n'importe lequel des gènes C_λ , chacun d'entre eux ayant la même structure bipartite J-C. En raison de cette structure multipartite, nous essaierons d'éviter la confusion sur la nature des séquences d'ADN en associant à V, J ou C les « segments » du génome ou des « régions » du polypeptide.

Une chaîne légère kappa est également assemblée à partir de deux morceaux, mais il existe une différence dans la structure de la région C. Un groupe de cinq segments J s'étale sur une région de 500-700 pb, séparée par un intron de 2-3 kb du seul exon C_κ . Chez la souris, le segment J central est non fonctionnel ($\psi J3$). Un segment V_κ peut être relié à n'importe lequel des segments J. Les résultats de la réaction de jonction kappa sont montrés dans la **Figure 32.5**.

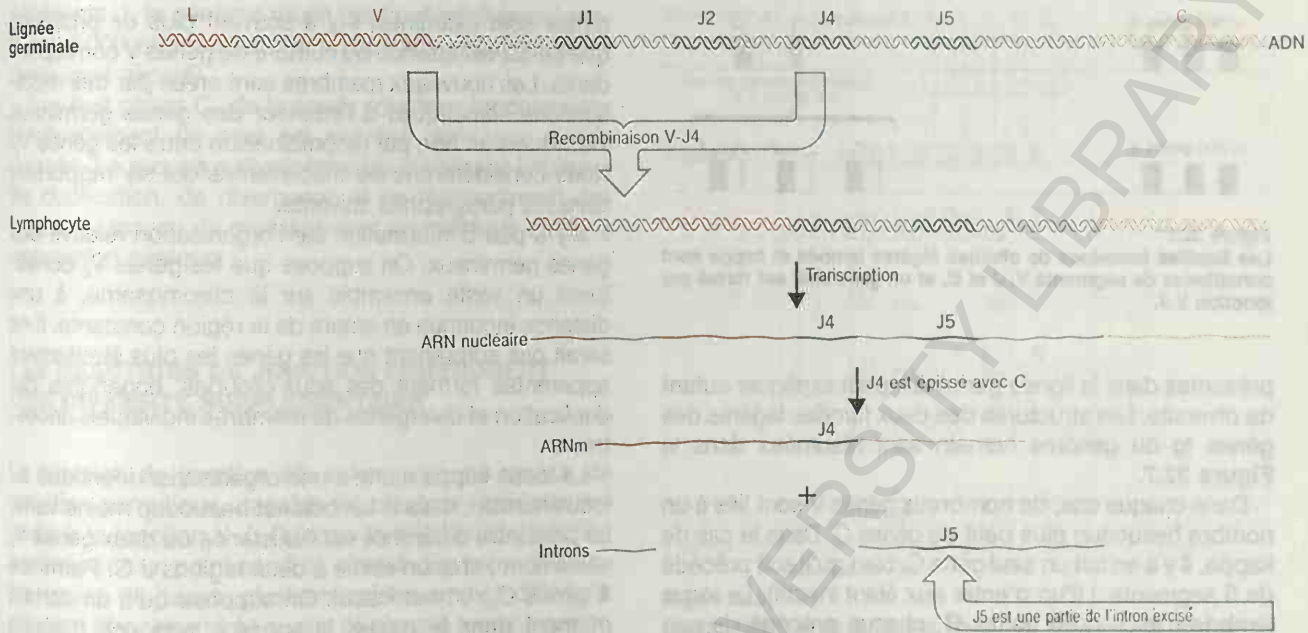
Quel que soit le segment J utilisé il devient la partie terminale de l'exon variable intact. Tous les segments J situés à gauche du segment J recombinant sont perdus

(J1 et J2 ont été perdus dans la figure). Tout segment J situé à droite du segment J recombinant est traité comme une partie de l'intron qui sépare les exons variable et constant (dans la figure, J5 est inclus dans la région excisée avant épissage des exons).

Tous les segments J fonctionnels possèdent un signal à leur extrémité gauche qui leur permet de se recombiner avec le segment V ; et il possèdent un signal à leur extrémité droite qui peut être utilisé pour le raccordement à l'exon C. Quel que soit le segment J reconnu dans la recombinaison de l'ADN, il utilise son signal d'épissage dans le remaniement de l'ARN.

Un segment supplémentaire est impliqué dans la construction d'un gène de chaîne lourde, comme il est montré dans la **Figure 32.6**. Le segment **D** (pour diversité) a été découvert par la présence dans la protéine de quelques acides aminés (2-13) entre les séquences codées par le segment V et le segment J. Un groupe de segments D se trouve sur le chromosome entre les segments V_H et les segments J_H . Un segment V_H se recombine avec un des segments D ; et le segment D se recombine avec un des quatre segments J_H (qui varient dans leur longueur entre 4 à 6 codons). La reconstruction conduit à l'expression du segment C_H adjacent (constitué de plusieurs exons).

L'organisation des segments D n'est pas totalement

**Figure 32.5**

Le gène C kappa est précédé de multiples segments J dans la lignée germinale ; la jonction V-J peut reconnaître n'importe quel segment J, qui est ensuite épissé avec la région C pendant la maturation de l'ARN.

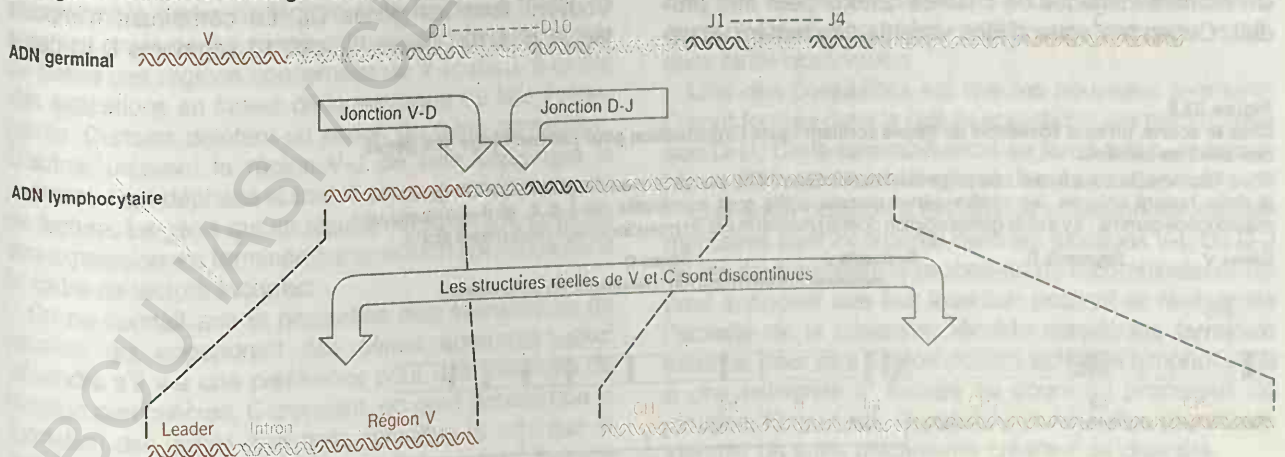
éclaircie, mais certains semblent se disposer en tandem. Le locus chaîne lourde de la souris contient environ dix segments D de longueur variable. Un mécanisme inconnu doit assurer que le *même* segment D soit impliqué dans les réactions de jonction V-D et D-J.

LA DIVERSITÉ DE L'INFORMATION GERMINALE

Maintenant on doit examiner les différents types de gènes V et C pour voir comment la variété des régions codantes

Figure 32.6

Les gènes lourds sont assemblés par jonction d'un segment V avec un segment D et par jonction du segment D avec un des segments J précédant le gène C.



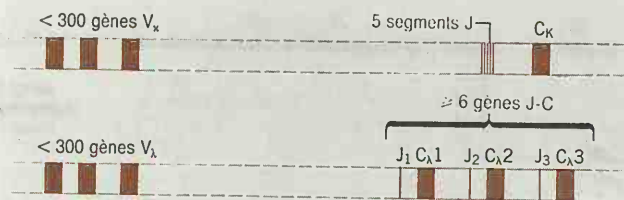


Figure 32.7

Les familles humaines de chaînes légères lambda et kappa sont constituées de segments V, J et C, et un gène actif est formé par jonction V-J.

présentes dans la lignée germinale peut expliquer autant de diversité. Les structures des deux familles légères des gènes Ig du génome humain sont résumées dans la Figure 32.7.

Dans chaque cas, de nombreux gènes V sont liés à un nombre beaucoup plus petit de gènes C. Dans le cas de kappa, il y a en fait un seul gène C, bien qu'il soit précédé de 5 segments J (l'un d'entre eux étant inactif). Le locus lambda a plus de 6 gènes C, chacun précédé de son propre segment J. Du fait du degré variable de divergence entre les gènes germinaux, il est difficile d'estimer le nombre de gènes V_κ et V_λ dans la lignée germinale. Une sonde individuelle peut réagir avec plusieurs gènes ; quand un de ces gènes est utilisé comme sonde, il peut réagir avec plusieurs autres ; et ainsi de suite. C'est la situation classique d'une famille de gènes répétitifs, dont certains membres ont une parenté assez proche et d'autres une parenté plus éloignée. Il y a probablement moins de 300 gènes V germinaux dans chaque famille.

Quelle est l'origine de la diversité de la région V ? En combinant n'importe lequel des 300 gènes V_κ avec n'importe lequel des 4 segments J, le génome murin peut potentiellement produire quelque 1 200 chaînes kappa. Un nombre identique de chaînes lambda peut être produit. Cependant, quand des variants étroitement appa-

rentés sont examinés, il y a souvent plus de protéines que ce qui est attendu du nombre de gènes V correspondants. Les nouveaux membres sont créés par des modifications somatiques à l'intérieur des gènes germinaux individuels (et non par recombinaison entre les gènes V). Nous considérerons les mécanismes qui s'y rapportent dans les paragraphes suivants.

Il y a peu d'information sur l'organisation relative des gènes germinaux. On suppose que les gènes V_κ constituent un vaste ensemble sur le chromosome, à une distance inconnue en amont de la région constante. Il ne serait pas surprenant que les gènes les plus étroitement apparentés forment des sous-groupes, engendrés par duplication et divergence de membres individuels ancestraux.

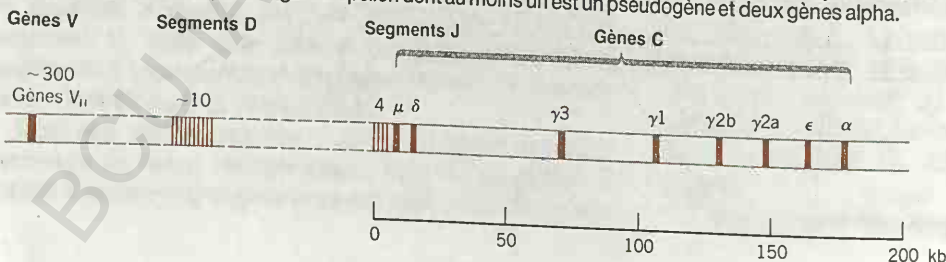
Le locus kappa murin a une organisation identique au locus humain, mais le lambda est beaucoup moins varié. La principale différence est l'existence de deux gènes V_λ seulement ; chacun est lié à deux régions J-C. Parmi les 4 gènes C_λ , un est inactif. On suppose qu'à un certain moment dans le passé, la souris a subi une délétion catastrophique de la plupart de ses gènes V_λ germinaux. Ceci peut expliquer pourquoi 95 % des cellules B murines expriment le type kappa de chaîne légère. Chez l'homme, environ 60 % des chaînes légères sont kappa et environ 40 % sont lambda. (Un lymphocyte donné synthétise soit une chaîne légère kappa, soit une chaîne légère lambda qui s'associera à la chaîne lourde.)

Le locus unique pour la production d'une chaîne lourde est constitué de plusieurs sections séparées, dont la structure dans le génome de souris est résumée dans la Figure 32.8. Il y a probablement quelque 300 gènes V_H . A une distance inconnue de ceux-ci se trouve le groupe des segments D. Et probablement pas très loin vient le groupe des segments J. Sur les 170 kb suivants se trouvent tous les gènes C_H . En combinant n'importe lequel des 100 gènes V_H , des 10 segments D et des 4

Figure 32.8

Chez la souris, un seul ensemble de gènes contient toute l'information pour l'assemblage des gènes des chaînes lourdes.

Chez l'homme, la structure a une organisation similaire, mais possède plus de gènes C_H . Les gènes mu et delta restent uniques, les quatre gènes gamma actifs sont numérotés de 1 à 4, et il y a aussi un pseudogène gamma ; il y a trois gènes epsilon dont au moins un est un pseudogène et deux gènes alpha.



segments J, le génome murin peut potentiellement produire 4 000 régions variables pour accompagner n'importe quel gène C_H.

Les huit gènes C_H de la souris sont tous fonctionnels ; habituellement un seul est exprimé dans une cellule donnée. Le groupe a divergé par les processus habituels de duplication, de divergence et de réassortiment des exons au moyen du crossing-over asymétrique ou de la conversion génique.

LES RÉACTIONS DE JONCTION ENGENDRENT UNE DIVERSITÉ SUPPLÉMENTAIRE

La réaction de jonction elle-même est une source de variation somatique, capable de changer l'acide aminé codé à la jonction V-J. Puisque la réaction de jonction se réalise en un point non fixe, c'est-à-dire entre différentes paires de bases, plusieurs acides aminés différents peuvent être engendrés au site de chaque recombinaison potentielle V-J. Un exemple est illustré dans la Figure 32.9 pour la jonction kappa.

L'utilisation de cinq cadres potentiels de recombinaison fait apparaître trois acides aminés différents à la position 96, dont un (arginine) *n'est pas codé dans la lignée germinale*. Puisque d'autres segments V_x et J_x ont des codons différents à ces positions, une grande diversité devient possible au point de jonction. Il est intéressant de savoir que l'acide aminé en position 96 forme une partie du site de liaison avec l'antigène et qu'il est impliqué dans les contacts entre les chaînes légères et les chaînes lourdes.

Une diversité encore plus grande peut être obtenue si la recombinaison n'est pas restreinte à des positions respectant le cadre de lecture et la longueur totale du segment, mais peut survenir entre n'importe quel couple de bases des régions concernées de V et de J. Il existe des indications en faveur de la survenue de tels événements. Certains délètent un codon du gène somatique. D'autres déphasent la région V-J de telle sorte que le segment J est déphasé et traduit dans un mauvais cadre de lecture. Le gène qui en résulte est aberrant, puisque son expression est terminée par un codon non-sens dans le cadre de lecture incorrect.

On ne connaît pas la proportion des événements de jonction qui engendrent des gènes aberrants ; ceci dépendra s'il y a une préférence pour des réactions de jonction particulières. Cependant, on peut penser que la formation des gènes aberrants constitue le prix que la cellule doit payer pour acquérir une plus grande diversité

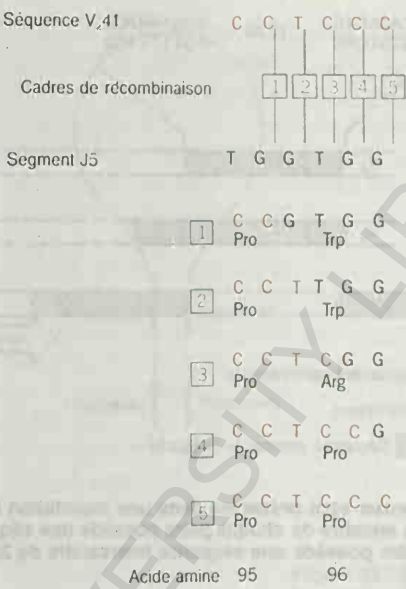


Figure 32.9 La jonction V-J de la chaîne kappa se fait dans les séquences codant pour les acides aminés 95 et 96. L'utilisation de sites alternatifs pour la recombinaison entre les segments V et J alignés crée de nouveaux codons. Des événements identiques se produisent au cours des réactions de jonction V-J des chaînes lambda et V-D-J des chaînes lourdes.

grâce à sa capacité d'ajuster le site de l'événement de recombinaison.

Une diversité similaire, bien qu'encore plus grande, est obtenue avec les réactions de jonction qui impliquent le segment D de la chaîne lourde. On retrouve le même problème de formation de gènes non productifs par des recombinaisons qui déphasent la lecture de J et C.

Il existe une énigme au sujet des segments D puisque certaines chaînes lourdes ont des résidus d'acides aminés qui ne correspondent à aucun des segments D connus. Deux hypothèses ont été proposées pour expliquer cette observation.

Une des possibilités est que les nouveaux segments D sont formés dans la cellule somatique par recombinaison D-D. Cette recombinaison se ferait selon un mécanisme différent de celui impliqué dans la jonction V-D-J.

Une autre possibilité est que des nucléotides supplémentaires sont incorporés dans les jonctions V-D ou D-J ou les deux pendant le processus de recombinaison. On peut supposer que leur insertion pourrait se réaliser via l'activité de la désoxynucléoside transférase terminale (connue pour être un composant actif des lymphocytes) à une extrémité 3' formée au cours du processus de jonction. En tout cas, l'existence de nouvelles séquences identifie un autre mécanisme créateur de diversité.

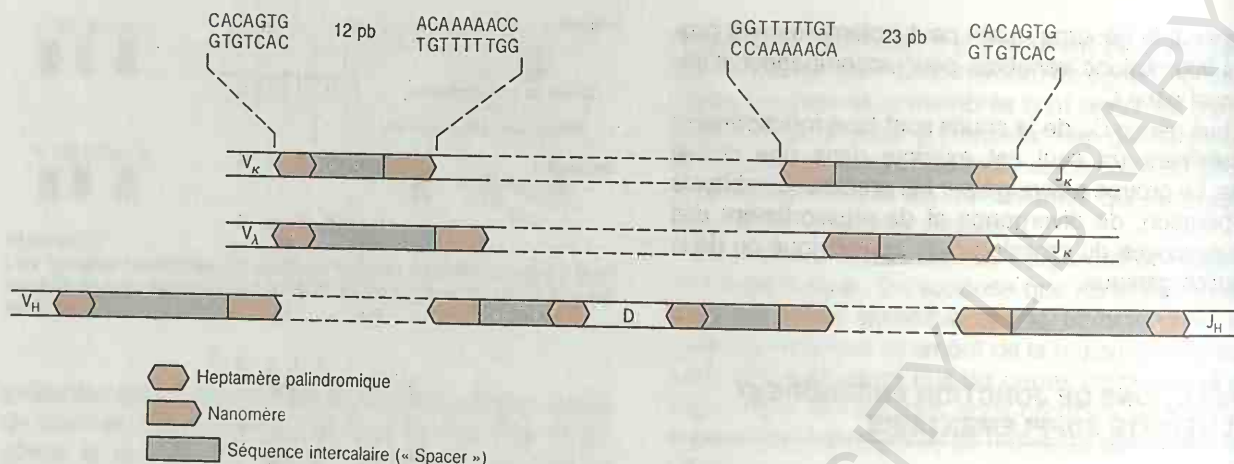


Figure 32.10

Les séquences consensus sont présentes dans une orientation inversée à chaque paire de sites de recombinaison. Un membre de chaque paire possède une séquence intercalaire de 12 pb entre ses composants ; l'autre possède une séquence intercalaire de 23 pb.

LA RECOMBINAISON DES GÈNES V ET C ENGENDRE DES DÉLÉTIONS ET DES RÉARRANGEMENTS

L'assemblage des gènes des chaînes légères et des chaînes lourdes implique le même mécanisme (bien que le nombre de morceaux soit différent). La présence des mêmes séquences consensus aux extrémités de tous les segments germinaux qui participent aux réactions de jonction indique l'emploi d'enzymes identiques ou très proches. Chaque séquence consensus est constituée d'un heptamère palindromique séparé par 12 ou 23 pb d'un nanomère.

La Figure 32.10 illustre la relation existant entre les séquences consensus situées aux loci Ig murins. Au locus kappa, chaque gène V_κ est suivi d'une séquence consensus ayant une séquence intercalaire (spacer) de 12 pb. Chaque segment J_κ est précédé d'une séquence consensus ayant une séquence intercalaire de 23 pb. Les séquences consensus V et J ont une orientation inversée. La disposition inverse est trouvée au locus lambda ; chaque gène V_λ est suivi d'une séquence consensus ayant une séquence intercalaire de 23 pb alors que chaque gène J_λ est précédé d'un consensus avec une séquence intercalaire de 12 pb.

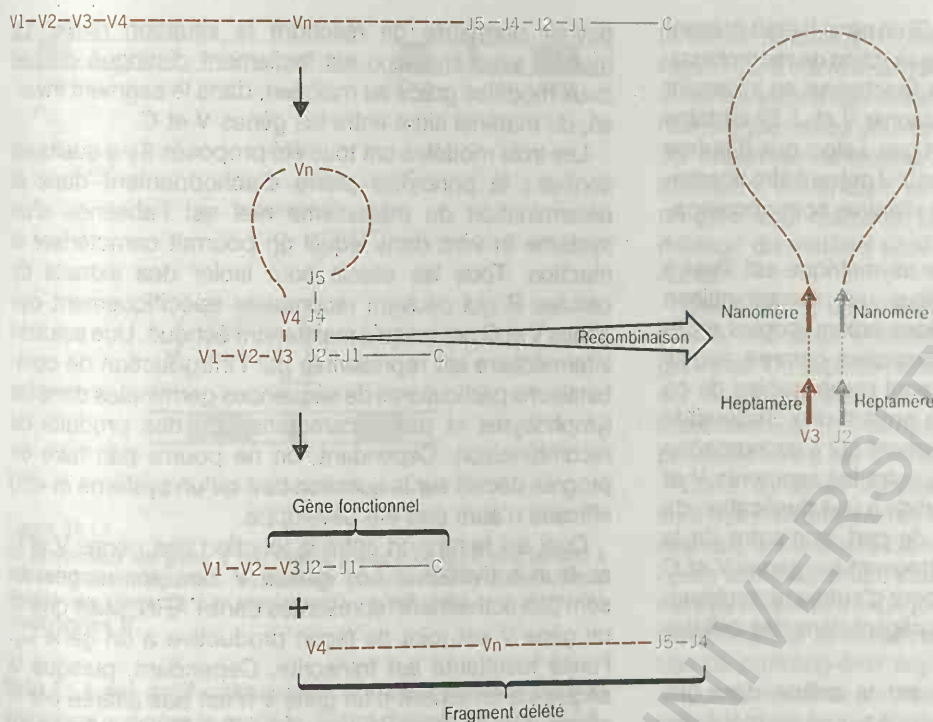
La comparaison de ces deux situations suggère une règle pour la réaction de jonction : une séquence consensus avec un type d'espacement ne peut être jointe qu'avec une séquence consensus ayant l'autre type

d'espacement. Puisque les séquences consensus situées sur les segments V et J peuvent se trouver dans l'une ou l'autre situation, les différents espacements ne fournissent aucune information directionnelle, mais ils servent à éviter qu'un gène V se recombine avec un autre ou qu'un segment J se recombine avec un autre.

Ce concept est confirmé par la structure des composants des gènes des chaînes lourdes. Chaque gène V_H est suivi d'une séquence consensus avec une séquence intercalaire de 23 pb. Les segments D sont flanqués sur chaque côté de séquences consensus avec une séquence intercalaire de 12 pb. Les segments J_H sont précédés de séquences consensus ayant une séquence intercalaire de 23 pb. Ainsi le gène V doit être uni à un segment D ; et le segment D doit être uni à un segment J. Un gène V ne peut pas être uni directement à un segment J, parce que les deux possèdent le même type de séquence consensus.

L'espacement entre les composants des séquences consensus correspond à environ un ou deux tours de la double hélice. Ce fait peut refléter une relation géométrique dans la réaction de recombinaison. Par exemple, la (les) protéine(s) de recombinaison peut (peuvent) approcher l'ADN d'un côté, de la même façon que l'ARN polymérase et les répresseurs approchent les éléments de reconnaissance tels que les promoteurs et les opérateurs (voir Chapitre 9).

L'utilisation du terme de « recombinaison » pour décrire la réunion des composants des gènes des immunoglobu-

**Figure 32.11**

Les séquences consensus pourraient être utilisées pour juxtaposer les régions V et J de telle sorte qu'une coupure et une réunion excisent le matériel situé entre elles.

lines ne signifie pas que la réaction implique une recombinaison réciproque entre séquences homologues. Ceci implique un réarrangement physique des séquences, mettant en jeu une rupture et une réunion (plutôt que des événements de type transposition) ; mais on ne connaît pas encore le mécanisme des événements moléculaires. Trois possibilités sont illustrées schématiquement dans les Figures 32.11 et 32.13.

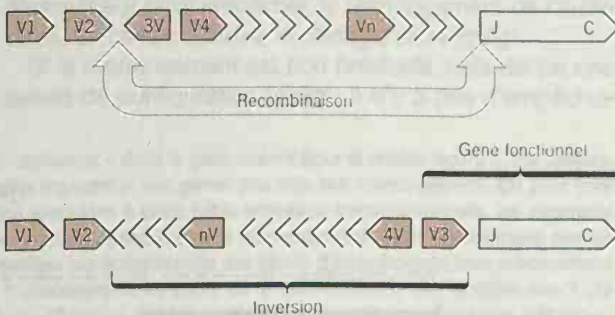
Le modèle par délétion est illustré dans la **Figure 32.11**. Il suppose qu'un des gènes V fusionne directement avec un des segments J. Tout le matériel situé entre les deux séquences réagissantes est excisé et vraisemblablement perdu. Les séquences réagissantes pourraient être réunies par appariement des séquences consensus inversées.

Notons qu'une recombinaison réciproque dans les séquences consensus inverserait la région située entre V et J au lieu de la déléter ; ainsi si l'appariement des séquences consensus conduit à une délétion, il doit s'agir d'un type différent de rupture et réunion. Une des hypothèses est que la boucle formée par l'appariement pourrait être enlevée par coupure à la base, avec ligature simultanée des extrémités des régions V et J.

Plusieurs modèles par inversion sont possibles ; l'un d'entre eux est illustré dans la **Figure 32.12**. Nous avons supposé jusqu'à maintenant que tous les gènes V ont la même orientation que les segments J bien que nous n'ayons pas encore d'information reliant les régions

Figure 32.12

Une réaction de jonction impliquant un gène V d'orientation opposée au segment J pourrait former un gène V-J-C fonctionnel en inversant la région comprise entre V et J.



codant pour V, aux segments J. Si un gène V était présent dans une orientation inverse, une réaction de recombinaison pourrait engendrer un gène fonctionnel en inversant le matériel intervenant pour fusionner V et J. Si certains gènes V ont la même orientation que J alors que d'autres sont inversés, certaines jonctions V-J doivent être accompagnées de délétions alors que d'autres sont accompagnées d'inversion.

Le modèle par crossing-over asymétrique est illustré dans la **Figure 32.13**. Il implique une recombinaison asymétrique entre les chromatides sœurs (copies répliquées) du gène. (Le modèle considère comme acquis que les séquences consensus sont responsables de ce phénomène.) La recombinaison produit une chromatide avec un gène fonctionnel, événement qui s'accompagne de la délétion du matériel situé entre les segments V et J recombinants. L'autre chromatide a une duplication de la région située entre V et J ; de part et d'autre de la région dupliquée aberrante se trouvent les gènes V et C qui peuvent rester disponibles pour d'autres recombinaisons. Les deux chromatides ségrègent dans des cellules filles, à la mitose.

La structure du gène actif est la même dans les modèles par délétion et par crossing-over asymétrique. Les modèles peuvent être distingués d'après le sort de la descendance de la cellule dans laquelle l'événement s'est produit. Dans le modèle avec délétion, tous les descendants ont la délétion. Dans le modèle avec crossing-over asymétrique, la moitié des descendants a la délétion, mais l'autre moitié a la duplication potentiellement réactive. L'analyse de la généalogie des cellules dans lesquelles la réaction de recombinaison se produit

devrait permettre de résoudre la situation réelle. Le modèle avec inversion est facilement distingué de ces deux modèles grâce au maintien, dans le segment inversé, du matériel situé entre les gènes V et C.

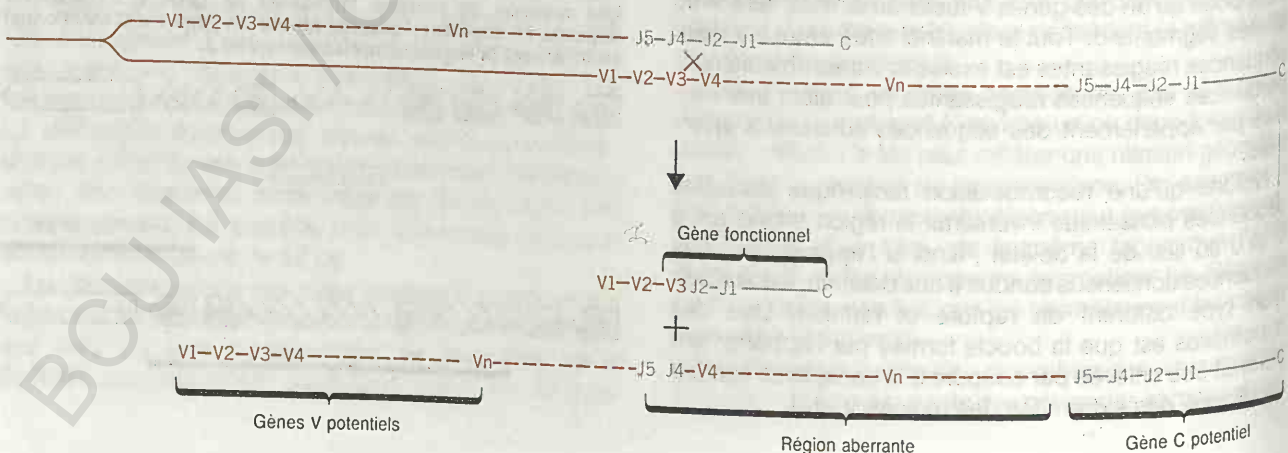
Les trois modèles ont tous été proposés il y a quelques années ; la principale pierre d'achoppement dans la détermination du mécanisme réel est l'absence d'un système *in vitro* dans lequel on pourrait caractériser la réaction. Tous les essais pour isoler des extraits de cellules B qui peuvent recombiner spécifiquement des gènes V et C ont jusqu'à maintenant échoué. Une solution intermédiaire est représentée par l'introduction de combinaisons particulières de séquences germinales dans les lymphocytes et par la caractérisation des produits de recombinaison. Cependant, on ne pourra pas faire de progrès décisifs sur la question tant qu'un système *in vitro* efficace n'aura pas été développé.

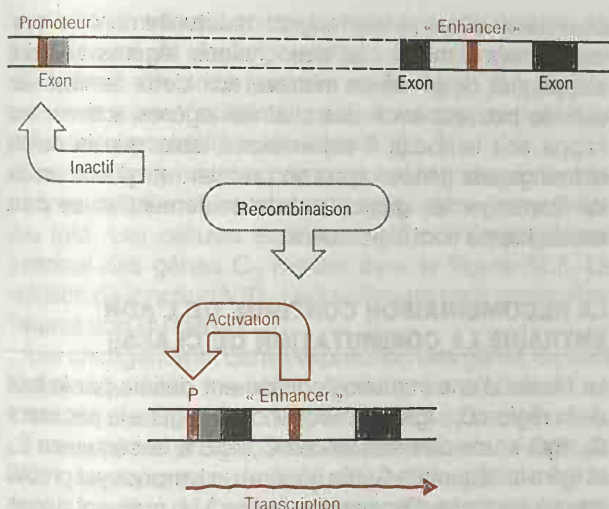
Quel est le rapport entre la jonction des gènes V et C et leur activation ? Les gènes V non réarrangés ne sont pas activement représentés dans l'ARN. Mais quand un gène V est joint de façon productive à un gène C_H, l'unité résultante est transcrite. Cependant, puisque la séquence en amont d'un gène V n'est pas altérée par la réaction de jonction, *le promoteur doit être le même pour les gènes non réarrangés, réarrangés de façon non productive, et réarrangés de façon productive.*

Si le promoteur V est activé par sa relocalisation au niveau de la région C, l'effet doit dépendre des séquences en aval. Quel rôle peuvent-elles jouer ? Le cas le mieux caractérisé est celui du locus lourd, où une séquence activatrice (enhancer) est présente dans le principal intron du premier gène C_H. Le « enhancer » est spécifique du

Figure 32.13

Un crossing-over asymétrique entre chromatides sœurs pourrait créer un recombinant avec un gène fonctionnel et une délétion de matériel entre V et J, l'autre recombinant ayant une duplication de la région située entre V et J.



**Figure 32.14**

Un promoteur de gène V est inactif jusqu'à ce qu'il soit amené à portée d'une séquence activatrice (« enhancer ») placée dans l'intron du gène C. Le « enhancer », n'est actif que dans les lymphocytes B.

tissu * ; il est actif uniquement dans les cellules B. Son existence suggère le modèle illustré dans la **Figure 32.14**, dans lequel le promoteur du gène V est activé dès qu'il est amené à proximité du « enhancer ». Il est vraisemblable que le même mécanisme est utilisé dans les autres loci des immunoglobulines.

QUELQUES CAUSES POSSIBLES D'EXCLUSION ALLÉLIQUE

Chaque cellule B exprime une combinaison particulière de chaînes d'immunoglobuline. Chaque recombinaison V-J ou V-D-J est différente. Un seul réarrangement productif de chaque type se produit dans un lymphocyte donné, pour produire un gène de chaîne légère et un gène de chaîne lourde. Puisque chaque événement implique que les gènes d'un seul des chromosomes homologues, *les allèles situés sur l'autre chromosome ne sont pas exprimés dans la même cellule*. Ce phénomène est appelé **exclusion allélique**.

L'exclusion allélique complique l'analyse de la recom-

binaison somatique. Une sonde réagissant avec une région qui a été réarrangée sur un chromosome détectera aussi les séquences alléliques situées sur le chromosome homologue. Nous pouvons ainsi être contraints d'analyser ensemble les destins différents des deux séquences alléliques. Cependant, le modèle habituel retrouvé dans un gène actif réarrangé peut être interprété comme une délétion du matériel situé entre les loci V et C recombinants.

Quelquefois, une cellule active montre, pour un gène donné, les deux modèles d'organisation, réarrangé et germinale. On suppose alors que la jonction s'est produite sur un chromosome, alors que l'autre chromosome est resté inaltéré. Le réarrangement pourrait s'être produit par délétion ou par crossing-over asymétrique.

Dans certaines lignées, les deux chromosomes ont subi des réarrangements indépendants. Dans certains de ces cas, le matériel situé entre les gènes V et C recombinants est entièrement absent de la lignée cellulaire. Ceci est très facilement expliqué par l'apparition d'événements indépendants au niveau de chaque chromosome (délétion ou crossing-over asymétrique).

Quand les deux chromosomes ont perdu leur configuration germinale, seul l'un d'entre eux a subi un **réarrangement productif** engendrant un gène fonctionnel. Le **réarrangement** sur l'autre chromosome a été **non productif** ; ceci peut prendre plusieurs formes, mais dans tous les cas la séquence du gène est aberrante et ne peut pas être exprimée sous forme d'une chaîne d'immunoglobuline.

La coexistence de réarrangements productifs et non productifs suggère l'existence d'une boucle de rétrocontrôle pour contrôler le processus de recombinaison. Supposons, comme le montre la **Figure 32.15**, que chaque cellule débute avec deux loci de configuration germinale non réarrangée Ig^0 . L'un ou l'autre de ces loci peut être réarrangé pour engendrer un gène productif Ig^+ ou un gène non productif Ig^- . Si le réarrangement est productif, la synthèse d'une chaîne active fournit un déclencheur pour empêcher le réarrangement de l'autre allèle. La cellule active a la configuration Ig^0/Ig^+ .

Si le réarrangement est non productif, cela donne une cellule de configuration Ig^0/Ig^- . Il n'y a pas d'empêche-

* N. d. T. : A noter que l'on n'avait toujours pas, en 1988, caractérisé de « enhancer » dans le gène codant pour la chaîne légère λ . Par ailleurs, il semble que, au cours de la différenciation des lymphocytes B, une certaine expression des gènes précède leur réarrangement. On peut faire l'hypothèse qu'à la faveur des modifications de la structure de la chromatine associées à cette faible activation transcriptionnelle, les segments d'ADN deviennent accessibles à une recombinaison (enzyme de recombinaison) qui catalyserait alors le réarrangement. Cette recombinaison semble identique dans les cellules B et dans les cellules T. C'est probablement l'activation transcriptionnelle des gènes d'immunoglobulines précocement au cours de la différenciation lymphocytaire B et des gènes du récepteur T précocement au cours de la différenciation des lymphocytes T qui explique que l'événement de recombinaison soit spécifique du type cellulaire. Plusieurs résultats utilisant des souris transgéniques militent en faveur de ce modèle.

ment pour le réarrangement de l'allèle germinal restant. Si ce réarrangement est productif, la cellule qui exprime a la configuration Ig^+/Ig^- . De nouveau, la présence d'une chaîne active supprime la possibilité de nouveaux réarrangements.

Deux réarrangements successifs non productifs produisent une cellule Ig^-/Ig^- . (En effet, le modèle stochastique prévoit que l'on devrait pouvoir trouver de telles cellules.) Cependant, il est possible que, dans certains cas, une cellule Ig^-/Ig^- puisse faire une nouvelle tentative. (Quelquefois les structures d'ADN observées peuvent avoir été engendrées uniquement par des réarrangements successifs.)

Le nœud du modèle est que la cellule continue à essayer de recombiner les gènes V et C jusqu'à ce qu'un réarrangement productif soit atteint. L'exclusion allélique est causée par la suppression d'un réarrangement ultérieur dès que la chaîne active est produite. L'utilisation de ce modèle *in vivo* est confirmée par des expériences où un gène reconstruit est introduit dans le génome murin. Les souris transgéniques qui expriment le gène supplémentaire suppriment le réarrangement des gènes endogènes*.

L'exclusion allélique est indépendante pour les loci des chaînes lourdes et des chaînes légères (les gènes des

chaînes lourdes se réarrangent habituellement en premier), mais dans le cas des chaînes légères elle doit s'appliquer de la même manière aux deux familles (les cellules peuvent avoir des chaînes légères actives, soit kappa soit lambda). Il est vraisemblable que la cellule réarrange ses gènes kappa en premier, et qu'elle essaie de réarranger les gènes lambda seulement si les deux essais kappa sont infructueux.

LA RECOMBINAISON CONTINUE DE L'ADN ENTRAÎNE LA COMMUTATION DE CLASSE

La classe d'une immunoglobuline est définie par le type de la région C_H . IgM a la séquence C_{μ} , IgD a la séquence C_{δ} , IgG a une des séquences C_{γ} , IgE a la séquence C_{ϵ} et IgA a la séquence C_{α} . En général un lymphocyte produit une seule classe d'immunoglobuline à un moment donné, mais la classe peut changer dans la descendance cellulaire. Ceci est accompli par une substitution du type de la région C_H qui est exprimé ; le changement d'expression est appelé **commutation de classe**.

La commutation implique le seul gène C_H ; le même gène V_H continue à être exprimé. Ainsi un gène V_H peut être exprimé successivement en combinaison avec plus d'un gène C_H . La même chaîne légère continue à être

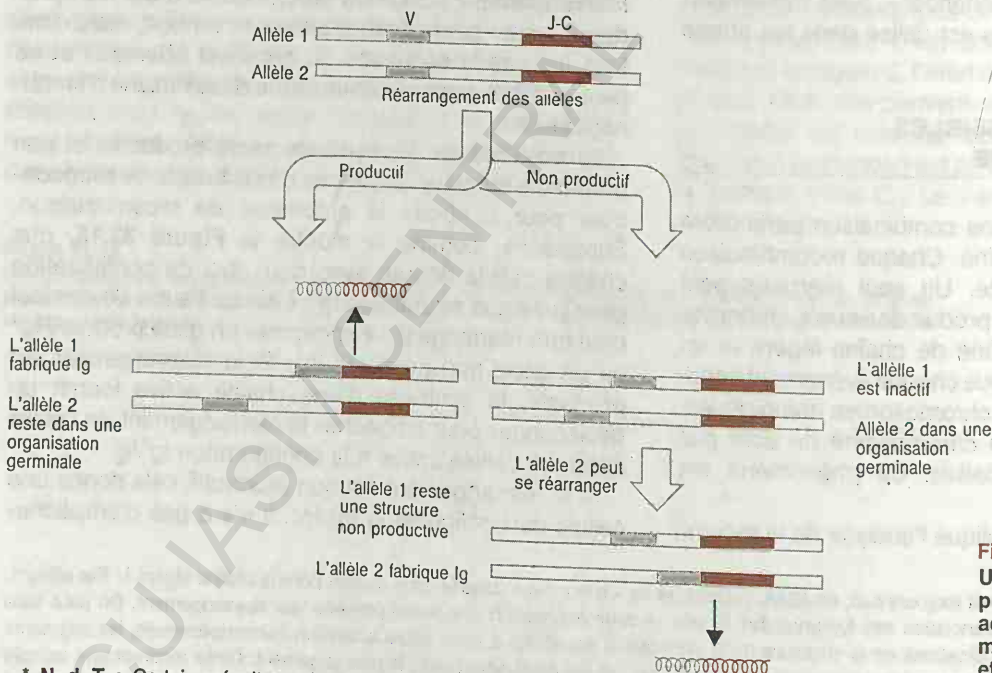


Figure 32.15

Un réarrangement fructueux, qui produit une chaîne légère ou lourde active, supprime tout réarrangement ultérieur de la même classe, et conduit à l'exclusion allélique.

* N. d. T. : Certains résultats très récents utilisant les souris transgéniques suggèrent que seule l'expression par un transgène réarrangé de la forme membranaire d'une chaîne d'immunoglobuline est capable d'entraîner l'inhibition du réarrangement des gènes endogènes. Un transgène modifié de telle sorte qu'il ne puisse diriger que la synthèse de la forme sécrétée de la chaîne d'immunoglobuline ne donne pas lieu au phénomène « d'exclusion ».

exprimée à travers toute la lignée de la cellule. La commutation de classe permet alors de changer de type de réponse effectrice (assurée par la région C_H), tout en conservant une facilité de reconnaissance antigénique constante (assurée par les régions V).

Tous les lymphocytes commencent leur vie productive comme cellules immatures engagées dans la synthèse des IgM. Les cellules exprimant IgM ont l'arrangement germlinal des gènes C_H montré dans la Figure 32.8. La réaction de jonction V-D-J est suffisante pour déclencher l'expression du gène C_H .

Les changements dans l'expression des gènes C_H sont réalisés de deux manières. Certains se produisent au cours de remaniement de l'ARN (voir section suivante). La majorité se produit via d'autres recombinaisons de l'ADN, impliquant un système différent de celui concerné par la jonction V-D-J (et capable d'opérer seulement ultérieurement).

Les cellules exprimant les gènes C_H tardifs ont habituellement des délétions des gènes C_μ et des autres gènes précédant le gène C_H exprimé. Ainsi la commutation de classe peut être accomplie grâce à une recombinaison qui amène un nouveau gène C_H en juxtaposition avec l'unité V-D-J exprimée. Les séquences des unités V-D-J- C_H commutées montrent que les sites de commutation résident en amont des gènes C_H eux-mêmes. Les sites de commutation sont appelés **régions S**. La Figure 32.16 représente deux commutations successives.

Au cours de la première commutation, l'expression de

C_{γ_1} succède à l'expression de C_μ . Le gène C_{γ_1} est amené dans la position exprimée par recombinaison entre les sites S_μ et S_{γ_1} , avec délétion du matériel situé entre les deux. Le site S_μ réside entre V-D-J et le gène C_μ . Le site S_{γ_1} réside en amont du gène C_{γ_1} . La région située entre V-D-J et le gène C_{γ_1} est retirée comme un intron pendant la maturation de l'ARN.

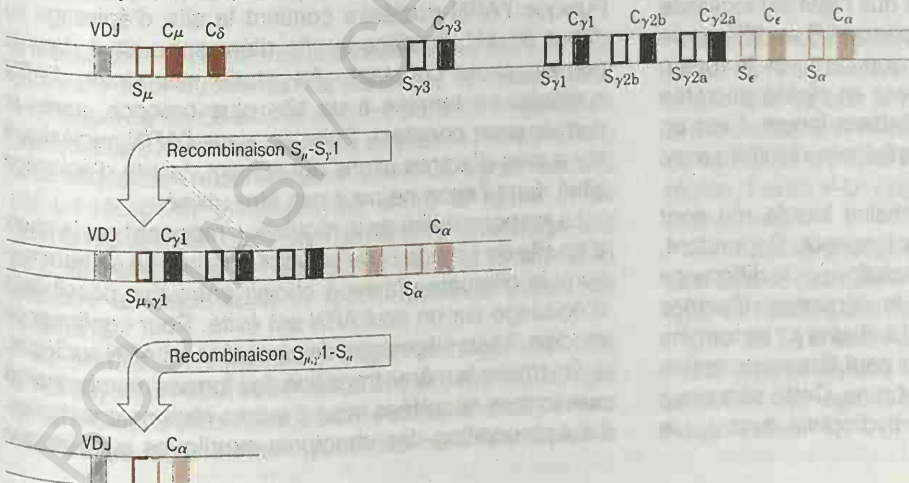
Le modèle avec délétion linéaire impose une restriction sur le locus du gène lourd : *dès qu'une commutation de classe a été réalisée, il devient impossible d'exprimer tout gène C_H qui résidait entre C_μ et le nouveau gène C_H* . Dans l'exemple de la Figure 32.16, les cellules exprimant C_{γ_1} devraient être incapables de donner naissance à des cellules exprimant C_{γ_3} qui a été délété.

Cependant, il devrait être possible, en principe, de réaliser une autre commutation sur n'importe quel gène C_H situé en aval du gène exprimé. La figure montre une seconde commutation permettant l'expression de C_α , réalisée par recombinaison entre S_α et la région de commutation S_{μ, γ_1} , formée par la première commutation.

On suppose que tous les gènes C_H possèdent des régions S en amont des séquences codantes. Mais on ignore s'il existe des restrictions à l'utilisation de ces régions S. Des commutations séquentielles se produisent en effet, mais on ne sait pas si elles constituent un moyen optionnel ou obligatoire pour aller jusqu'aux derniers gènes C_H . On aimerait savoir si IgM peut commuter directement avec toute autre classe.

On sait que les sites de commutation ne sont pas

Figure 32.16
La commutation de classe des gènes lourds peut se produire par recombinaison entre les régions commutatrices (S) en délétant le matériel situé entre les sites S recombinants. Des commutations successives peuvent se produire.



précisément définis, parce qu'il se trouve que des cellules différentes exprimant le même gène C_H ont recombinaison en des points différents. Quand on aura séquencé assez de ces sites S, on pourra définir les limites des régions à l'intérieur desquelles la commutation peut se produire. En attendant, nous pouvons rechercher au niveau des séquences de recombinaison, des caractéristiques communes qui pourraient jouer un rôle.

Les régions S se situent à environ 2 kb en amont des gènes C_H . Trois sites S_H distincts ont été caractérisés ; deux d'entre eux montrent une homologie, mais pas l'autre. Il y a des homologies avec certaines des régions S des autres gènes, mais pas de toutes. Des répétitions en tandem de courtes séquences sont présentes dans plusieurs régions S, mais on doit encore élucider leur signification.

Comme pour la jonction des gènes V et C, on sait que des délétions se produisent pendant la commutation C_H ; mais ceci ne prouve pas que le mécanisme sous-jacent est une délétion moléculaire ; la réaction pourrait s'effectuer par des échanges entre chromatides sœurs.

L'EXPRESSION PRÉCOCE DES CHAÎNES LOURDES PEUT ÊTRE MODIFIÉE PAR MATURATION DE L'ARN

La période de synthèse des IgM qui marque le début du développement lymphocytaire se divise en deux parties, pendant lesquelles des versions différentes de la région constante μ sont synthétisées. Lorsqu'une cellule souche se différencie en un lymphocyte pré-B, une chaîne légère est synthétisée, et la molécule d'IgM ($L_2\mu_2$) apparaît à la surface de la cellule. Cette forme contient la version μ_m de la région constante (m indique que l'IgM est localisée dans la membrane). Quand le lymphocyte B se différencie ultérieurement en plasmocyte, la version μ_s de la région constante est exprimée. (L'IgM est en réalité sécrétée sous forme de pentamère IgM₅J, dans lequel J est un peptide de jonction qui forme des liaisons disulfure avec les chaînes μ).

Les versions μ_m et μ_s de la chaîne lourde mu sont identiques sur la majorité de leur longueur. Cependant, la chaîne μ_m est plus longue que la chaîne μ_s ; la différence résulte d'un changement dans la séquence d'acides aminés à l'extrémité C-terminale. La chaîne μ_m se termine par une séquence hydrophobe qui peut être responsable de son implantation dans la membrane. Cette séquence est remplacée par une séquence hydrophile dans μ_s ; la

substitution permet à la chaîne lourde mu de passer à travers la membrane.

Les chaînes μ_m et μ_s sont codées par des ARNm différents. Les deux ARNm partagent une séquence identique jusqu'à l'extrémité du dernier domaine constant. Ensuite ils diffèrent ; μ_m a 41 codons de plus suivis par une queue non traduite ; et μ_s a 20 codons suivis par une queue différente non traduite.

La séquence génomique montre que les régions terminales de μ_m et μ_s sont codées par des exons différents. La relation entre la structure du gène et les ARNm est illustrée dans la **Figure 32.17**.

Quand l'immunoglobuline est liée à la membrane, la région constante est produite après épissage de six exons. Les quatre premiers codent pour les quatre domaines de la région constante. Les deux derniers, M1 et M2, codent pour les 41 résidus de la région C-terminale hydrophobe et pour la queue non traduite.

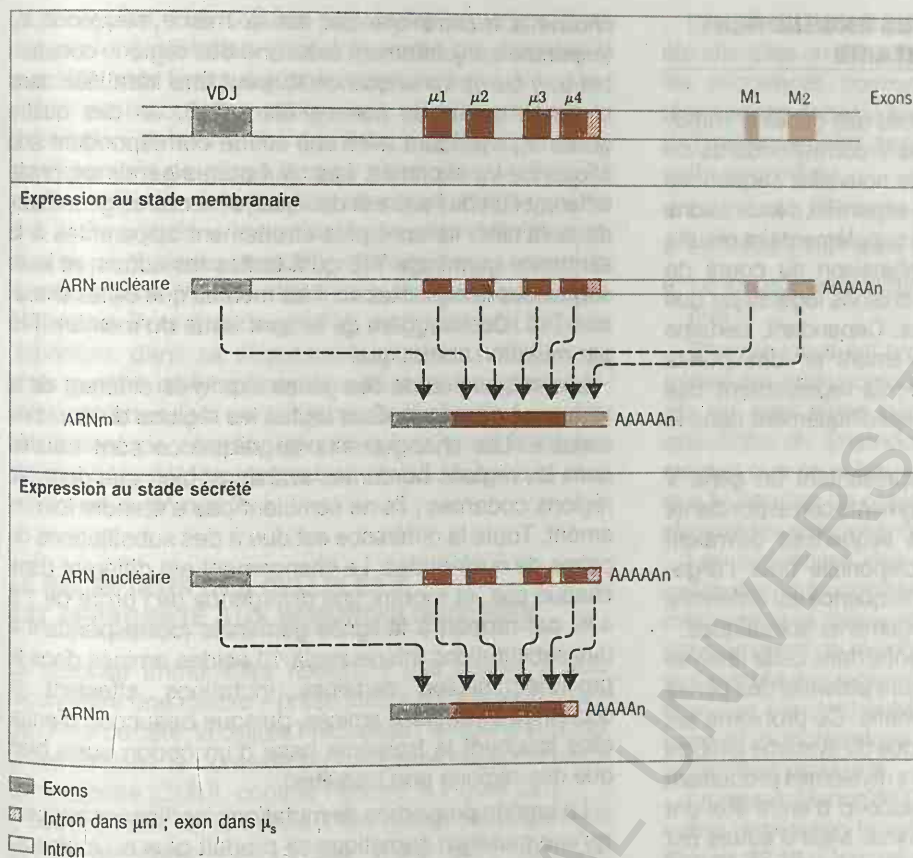
Au stade sécrétoire, la région constante est formée par jonction des quatre premiers exons seulement. Le dernier de ces exons s'étend au-delà de son extrémité dans le transcrit μ_m ; il comporte les 20 derniers codons et la queue non traduite.

Ces deux ARNm μ_s et μ_m sont ainsi alternativement produits selon qu'un site d'épissage situé dans le 4^e exon (codant pour le dernier domaine constant) est épissé aux exons M1 et M2 (avec raccourcissement de l'exon et formation de la séquence codant pour μ_m) ou est ignoré (avec allongement de l'exon et formation de la séquence codant pour μ_s). Comment l'utilisation de ce site d'épissage est-elle contrôlée ?

Un modèle plausible est illustré dans la **Figure 32.17**. Au stade membranaire, l'ARN nucléaire a une extrémité polyadénylée située après M2 (extrémité créée par clivage endonucléolytique ou bien peut-être par terminaison). Puisque l'ARN nucléaire contient le site d'épissage au début de M1, il utilise le site d'épissage situé dans le dernier exon constant. Au stade sécrétoire, l'ARN nucléaire se termine à un site plus précoce, après le dernier exon constant. Puisque, dans l'ARN nucléaire, il n'y a plus d'autres exons qui suivent, le site d'épissage situé dans l'exon ne peut pas être utilisé.

La responsabilité de la régulation reposant sur le choix d'un site de clivage et de polyadénylation (ou de terminaison), le dilemme d'avoir à choisir entre des possibilités d'épissage sur un seul ARN est évité. Pour confirmer ce modèle, il sera nécessaire de caractériser l'ARN nucléaire.

On trouve la même transition des formes membranaires aux formes sécrétées avec d'autres régions constantes. La conservation des structures exoniques suggère que

**Figure 32.17**

Le site de terminaison ou de clivage et de polyadénylation peut contrôler l'utilisation des sites d'épissage de telle sorte que des formes alternatives du gène lourd sont exprimées.

le mécanisme est le même. Les exons M1 et M2 des trois chaînes γ sont étroitement apparentés, et codent pour une séquence de 66 acides aminés, dont les 41 premiers montrent une homologie avec la séquence membranaire μ . La séquence conservée peut fournir un signal reconnu par un récepteur protéique qui ancre la forme m de l'immunoglobuline dans la membrane. L'ancrage est peut-être associé à la nécessité d'initier la prolifération cellulaire en réponse à la reconnaissance initiale de l'antigène.

La production simultanée d'IgM et d'IgD dans les lymphocytes B matures constitue une exception à la règle qu'un seul type d'immunoglobuline est synthétisé par une cellule. Les deux immunoglobulines sont identiques à

l'exception des régions constantes μ et δ de la chaîne lourde qui se substituent l'une à l'autre. Il semble vraisemblable que ce soit le résultat de voies alternatives pour la maturation de l'ARN. La région constante δ est proche de la région μ ; si la transcription continue sur toute la région, l'exon V-D-J pourra être épissé avec la série des exons delta.

Cette situation rappelle celle impliquée dans l'expression tardive des adénovirus, et on peut invoquer le même type de modèle que celui illustré précédemment dans la Figure 18.7 ; là encore, la formation de différentes extrémités polyadénylées dans l'ARN nucléaire contrôle la sélection des régions codantes qui seront épissées avec l'exon initial.

DES MUTATIONS SOMATIQUES ENGENDRENT UNE DIVERSITÉ SUPPLÉMENTAIRE

La comparaison entre les séquences des gènes d'immunoglobuline exprimés et les gènes V correspondants de la lignée germinale montre que de nouvelles séquences apparaissent dans la population exprimée. Nous avons vu qu'une partie de cette diversité supplémentaire résulte de décalages du site de recombinaison au cours de l'assemblage des exons V-J des chaînes légères ou des exons V-D-J des chaînes lourdes. Cependant, certains changements se produisent en amont en des points situés dans le domaine variable ; ils représentent des **mutations somatiques** induites spécifiquement dans le lymphocyte actif.

On peut utiliser une sonde représentant un gène V exprimé pour identifier tous les fragments correspondants dans la lignée germinale. Leurs séquences devraient identifier le répertoire complet disponible pour l'organisme. Tout gène exprimé dont la séquence est différente a dû être engendré par des changements somatiques.

Le principal problème expérimental dans cette analyse est de s'assurer que chaque membre potentiel des gènes V germinaux a réellement été identifié. Ce problème est surmonté par la simplicité inattendue du système lambda de la souris. Une étude de plusieurs myélomes produisant des chaînes λ_1 a montré que beaucoup d'entre eux ont la séquence du gène unique germinal. Mais d'autres ont de nouvelles séquences qui peuvent avoir été engendrées par mutation du gène terminal.

Afin de déterminer la fréquence d'une mutation somatique dans d'autres cas, il est nécessaire d'étudier un grand nombre de cellules dans lesquelles le même gène V est exprimé. Un procédé pratique pour identifier un tel groupe consiste à caractériser les immunoglobulines d'une série de cellules, qui expriment toutes une réponse immunitaire vis-à-vis d'un antigène spécifique.

(Les antigènes utilisés dans ce but sont des petites molécules dont la structure discrète va vraisemblablement provoquer une réponse uniforme, contrairement à une grosse protéine dont les différentes parties peuvent induire différents anticorps. Les petites molécules sont appelées des haptènes et sont conjuguées à une protéine non réactive pour former l'antigène. Les cellules sont obtenues en immunisant des souris avec l'antigène, en obtenant les lymphocytes réactifs, et quelquefois en fusionnant ces lymphocytes avec une cellule myélomateuse pour former un **hybridome** qui continue à exprimer l'anticorps désiré indéfiniment.)

Une étude de 19 lignées cellulaires différentes produisant des anticorps dirigés contre l'haptène phosphoryl-

choline a montré que dix ont la même séquence V_H (exprimée conjointement avec une des régions constantes μ , γ ou α). La séquence V_H peut être identifiée dans la lignée germinale comme étant T15, un des quatre gènes V_H hybridant avec une sonde correspondant à la séquence V_H exprimée. Les neuf autres gènes exprimés diffèrent l'un de l'autre et des quatre membres germinaux de la famille. Ils sont plus étroitement apparentés à la séquence germinale T15 qu'à toutes les autres, et leurs séquences flanquantes sont les mêmes que celles entourant T15. Ceci suggère qu'ils sont issus du membre T15 par mutation somatique.

Les séquences de ces gènes exprimés diffèrent de la séquence germinale dans toutes les régions du domaine variable. Les changements de séquence sont trouvés dans les régions bordantes aval aussi bien que dans les régions codantes ; ils ne semblent pas s'étendre loin en amont. Toute la différence est due à des substitutions de paires de nucléotides. Le changement est différent dans chaque cas, et montre une divergence de l'ordre de 1 à 4 % par rapport à la lignée germinale (correspondant à des substitutions inférieures à 10 acides aminés dans la protéine ; seules certaines mutations affectent la séquence des acides aminés, puisque beaucoup d'entre elles touchent la troisième base d'un codon aussi bien que des régions non traduites).

La grande proportion de mutations inefficaces suggère qu'une mutation somatique se produit plus ou moins au hasard dans une région incluant le gène V et s'étendant au-delà de celui-ci. Dans les gènes des chaînes lourdes, la commutation de classe peut fournir le déclenchement pour une mutation. Les gènes V_H associés aux chaînes constantes μ sont rarement mutés ; les (mêmes) gènes V_H associés aux chaînes constantes γ ou α peuvent être mutés.

On connaît maintenant plusieurs cas où une seule famille de gènes V est utilisée de façon constante pour répondre à un antigène spécifique. Le besoin d'une famille spécifique donnée nous renvoie au dilemme de savoir comment l'organisme fournit une réponse à une gamme imprévisible d'antigènes.

La spécificité de la réponse démontre l'importance d'une grande diversité du potentiel codant des gènes V. Après exposition à un antigène, il est vraisemblable que la région V, ayant l'affinité intrinsèque la plus élevée fournisse un point de départ. Les mutations aléatoires ont des effets imprévisibles sur la fonction protéique ; certaines peuvent inactiver la protéine, d'autres peuvent lui conférer une spécificité élevée vis-à-vis d'un antigène spécifique. Ainsi une des caractéristiques critiques du processus pourrait être la sélection, parmi la population

lymphocytaire, des cellules portant les anticorps dans lesquels une mutation aléatoire a créé un domaine V favorable pour lier tout antigène qui se trouve à portée.

Ceci ne résout pas entièrement un dilemme sur la constitution de la diversité des anticorps. Considérons le grand nombre de gènes V, sans mentionner la diversité additionnelle due à la jonction alternative des segments D et J. On voit dans le cas de la phosphorylcholine que, dans de nombreuses stimulations antigéniques séparées, le même gène V_H est utilisé pour donner naissance à l'anticorps. Pour que ce gène V_H particulier soit toujours disponible, dans sa séquence originale ou mutée, une vaste gamme de lymphocytes a dû subir des réarrangements en attendant un antigène. Ceci signifie-t-il qu'il y a tellement de réarrangements qui se produisent que chaque gène V potentiel est en réalité disponible sous une forme exprimée ?

LE RÉCEPTEUR DES CELLULES T EST APPARENTÉ AUX IMMUNOGLOBULINES

La réponse immunitaire nécessite une cellule T pour reconnaître une cellule « présentant » un antigène étranger. Pour ce faire, la cellule T reconnaît l'antigène étranger présenté dans le contexte du complexe majeur d'histocompatibilité (CMH), comme l'illustre la Figure 32.1.

Le récepteur de la cellule T est responsable de cette double reconnaissance. Il doit, d'une part, de façon analogue aux immunoglobulines, reconnaître un antigène étranger de structure imprévisible, et, d'autre part, reconnaître le système majeur d'histocompatibilité. On a pensé que la nature pouvait résoudre le problème de la reconnaissance antigénique par les cellules B et les cellules T de la même façon, et que dans ce cas on pouvait s'attendre à ce que l'organisation du récepteur des cellules T ressemble à celle des immunoglobulines avec des régions variables et des régions constantes.

Le récepteur est une glycoprotéine d'environ 80 000 daltons, constituée d'une chaîne α et d'une chaîne β , chacune d'environ 40 000 daltons *. Les chaînes sont maintenues ensemble par des liaisons disulfure. La cartographie peptidique suggère que les récepteurs des différentes lignées cellulaires T ont une partie de leurs séquences commune, mais diffèrent pour d'autres parties (comme les immunoglobulines). Le récepteur est une protéine de surface, probablement ancrée dans la membrane.

* N. d. T. : Le récepteur T de formule $\alpha\beta$ constitue en effet l'espèce prédominante, responsable de la spécificité des lymphocytes T matures. Il existe une autre molécule de récepteur formée des chaînes γ et δ , chacune étant codée par des gènes différents qui se réarrangent selon les mêmes modèles que les gènes α et β . Le gène δ se trouve à l'intérieur du locus α , dans des régions séparant les différents segments. Le récepteur $\gamma\delta$ semble caractériser des formes immatures de lymphocytes T, ainsi que des populations épithéliales et épidermiques de cellules T.

Les gènes codant pour le récepteur des cellules T ont été identifiés en découvrant des ARNm de cellules T dont les séquences correspondantes du génome ont été réarrangées dans les cellules T mais pas dans les cellules B. Trois ensembles de gènes ont été identifiés en utilisant cette approche :

- Le locus α TcR code pour la chaîne α du récepteur.
- Le locus β TcR code pour la chaîne β du récepteur.
- Le locus γ code pour une protéine de fonction inconnue *.

Chaque locus est organisé de la même façon que les gènes des immunoglobulines, avec des segments séparés qui sont réunis par une réaction de recombinaison spécifique du lymphocyte T. Les composants sont les mêmes que ceux trouvés dans les trois familles Ig, bien que le nombre de copies soit inférieur. Le Tableau 32.2 résume les composants des trois loci.

Les séquences des gènes TcR sont apparentées à celles des immunoglobulines. Les séquences V peuvent même avoir la même alternance de régions relativement constantes et de régions hypervariables. La région C est apparentée aux régions Ig constantes et possède un seul domaine constant suivi par des portions transmembranaire et cytoplasmique. La structure exon-intron reflète la fonction protéique.

La ressemblance de l'organisation des gènes TcR avec celles des gènes Ig est frappante, comme le montre la Figure 32.18. L'organisation de TcR α ressemble à celle de Ig α , avec des gènes V séparés d'un ensemble de segments J qui précède un seul gène C. Les composants de TcR β ressemblent à ceux de IgH ; l'organisation est différente, avec des gènes V séparés de deux ensembles contenant chacun un segment D, plusieurs segments J

Tableau 32.2
Trois loci avec des composants ressemblant aux segments des immunoglobulines sont réarrangés dans les lymphocytes T.

Locus	V	J	D	C
α	~ 40	~ 55	-	1
β	~ 50	12	2	2
γ	~ 5	3	-	3

α et β codent pour les composants du récepteur cellulaire T, mais la fonction de γ est inconnue. Le nombre de copies de chaque composant montré dans le tableau est similaire chez l'homme et la souris.

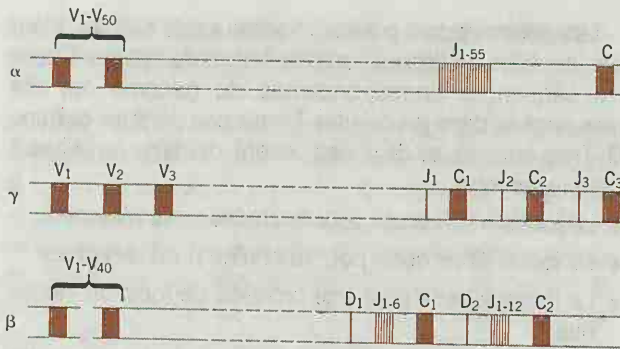


Figure 32.18

Trois loci qui se réarrangent spécifiquement dans les lymphocytes T ont une organisation gémale apparentée à celle des gènes des immunoglobulines.

et un gène C. L'organisation du locus γ ressemble à celle de Ig λ, avec des gènes V séparés d'une série de segments J, chacun étant lié à son propre gène C.

Les mêmes mécanismes doivent être vraisemblablement impliqués dans les réactions qui recombinaient les gènes Ig des cellules B et les gènes TcR des cellules T. Les segments TcR recombinants sont encadrés par des séquences consensus de type nanomère et heptamère identiques à celles utilisées par les gènes Ig. Ceci suggère fortement que les mêmes enzymes sont impliquées ; comment le processus est-il contrôlé pour que les loci Ig soient réarrangés dans les cellules B pendant que les récepteurs T sont réarrangés dans les cellules T ? (Voir N. du T. page 653).

Des modèles de reconstruction génique impliquant les séquences consensus sont illustrés dans les Figures 32.11 à 32.13. Une réaction d'inversion analogue à celle de la Figure 32.12 se produit dans au moins un cas où un segment TcR V_β a été identifié 10 kb en aval des gènes C_β . La Figure montre une réaction d'inversion pour un segment V en amont des segments J-C ; puisque le segment TcR V est en aval des segments D-J-C, dans ce cas précis ce doit être la région D-J-C qui est inversée.

Le lignage lymphocytaire présente un exemple clair de l'opportunisme des processus évolutifs : un procédé identique est utilisé à la fois dans les cellules B et les cellules T pour former des protéines qui possèdent une région variable capable de fournir une diversité significative, alors que les régions C sont plus limitées et représentent une petite gamme de fonctions effectrices.

Il reste à résoudre un dilemme central au sujet de la

fonction des cellules T. Les cellules T doivent s'engager dans deux processus de reconnaissance. Premièrement, elles doivent reconnaître l'antigène présenté par une molécule du complexe majeur d'histocompatibilité. Deuxièmement, elles doivent reconnaître simultanément cette molécule d'histocompatibilité elle-même, l'antigène n'étant reconnu par un lymphocyte T que lorsqu'il est présenté par les propres molécules d'histocompatibilité de l'organisme auquel appartient ce lymphocyte. Il existe plusieurs antigènes d'histocompatibilité différents, de telle sorte qu'une diversité considérable est nécessaire pour assurer la spécificité. Cependant, même en permettant l'introduction d'une variabilité supplémentaire au cours du processus de recombinaison TcR, il pourrait ne pas y avoir assez de versions différentes des récepteurs des cellules T disponibles pour assurer cette spécificité. Cela signifie-t-il qu'il existe un autre récepteur des cellules T à identifier ? Le gène γ pourrait-il coder pour une partie de ce second récepteur ? Ou le gène γ a-t-il quelque autre fonction ? *

COMPLEXITÉ DES PRINCIPAUX LOCI D'HISTOCOMPATIBILITÉ

Le principal locus d'histocompatibilité occupe un petit segment d'un seul chromosome chez la souris (où il est appelé locus *H2*) et chez l'homme (appelé locus *HLA*). Dans ce segment existent de nombreux gènes codant pour des fonctions impliquées dans la réponse immunitaire. Au niveau de ces loci géniques individuels dont les produits ont été identifiés, de nombreux allèles ont été trouvés dans la population ; le locus est décrit comme hautement **polymorphe** : ceci signifie que les génomes individuels doivent être vraisemblablement différents les uns des autres.

Les différents types de fonction identifiés dans *H2* et la région adjacente sont résumés sur la carte de la Figure 32.19. La région *H2* classique occupe 0,3 unité de recombinaison et inclut trois classes de gènes. La région adjacente recouvre une autre unité de recombinaison ; à l'intérieur de celle-ci, il y a des gènes apparentés à ceux présents dans la région *H2*. Ce « petit segment » de chromosome peut être mesuré en termes moléculaires. Les 1,3 unités de recombinaisons réunies représentent potentiellement plus de 2 500 kb d'ADN. La taille du locus humain est comparable ; il contient des fonctions identiques, bien que dans un ordre différent.

Les gènes de classe I codent pour les **antigènes de**

* N. d. T. : On sait maintenant que la combinatoire des différents segments V_α , J_α , d'une part, et V_β , D_β , J_β , d'autre part, est très probablement suffisante pour créer la diversité requise par la multiplicité des antigènes reconnus. Le second récepteur identifié, formé des produits des gènes γ et δ, n'a pas pour rôle d'augmenter la taille du répertoire des récepteurs T.

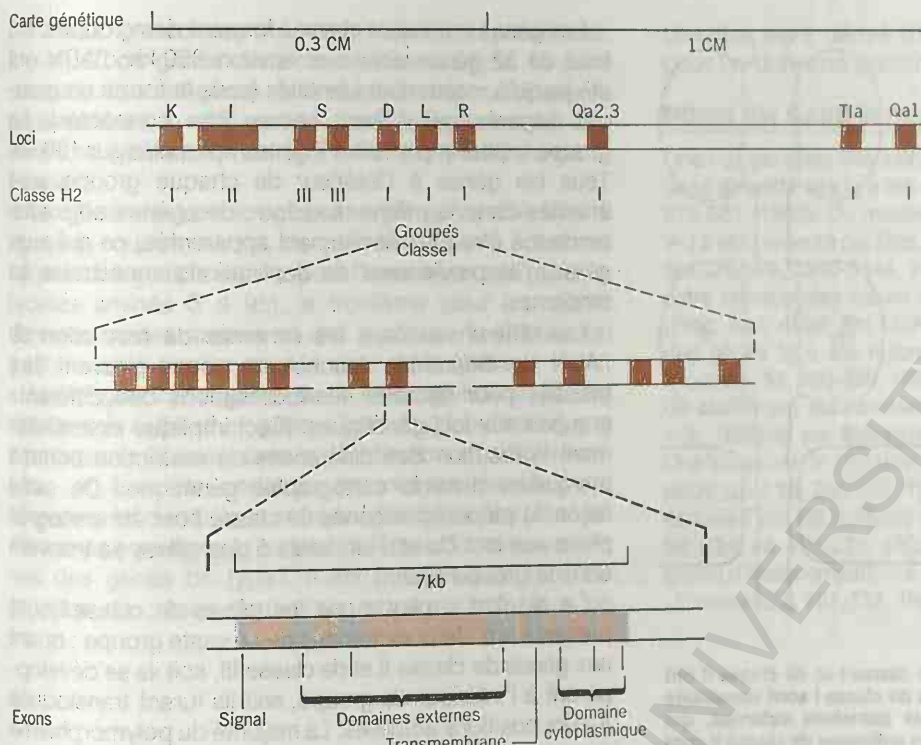


Figure 32.19

Le locus d'histocompatibilité de la souris contient plusieurs régions génétiques ; les gènes de classe I sont organisés en structures étendues ; la structure du gène individuel correspond à la structure de la protéine.

transplantation, protéines présentes sur chaque cellule de mammifère. Comme leur nom le suggère, ces protéines sont responsables du rejet d'un tissu étranger, qui est reconnu comme tel à cause de son arrangement particulier d'antigènes de transplantation. Dans le système immunitaire, leur présence sur certains lymphocytes T (cytotoxiques) joue un rôle dans la destruction des cellules cibles. Les régions K, D, L et R codent toutes pour des protéines de ce type ; ces fonctions sont définies sérologiquement (par leurs propriétés antigéniques). Chaque souche murine possède un des nombreux allèles possibles pour chacune de ces fonctions.

Les gènes de classe II codent pour des protéines trouvées à la surface des lymphocytes B et T et des macrophages. Ces protéines sont impliquées dans les communications intercellulaires qui sont nécessaires pour développer la réponse immunitaire. Connus au début comme étant la région I, les gènes de classe II sont maintenant divisés en sous-régions A, B, J, E et C (non montrées sur la carte)*.

Les gènes de classe III codent pour les **protéines du**

complément. Cette partie du locus est également connue comme étant la région S ; S signifie sérum, et indique que les protéines sont des composants du sérum. Leur rôle est d'interagir avec les complexes antigènes-anticorps pour provoquer la lyse des cellules.

À la droite de la région H2 se trouvent les loci Qa et Tla, qui codent pour des protéines localisées sur les cellules hématopoïétiques. Les protéines sont décrites comme des antigènes de différenciation, parce que chacune d'elles est trouvée seulement sur une sous-classe particulière des cellules sanguines, vraisemblablement en relation avec leur fonction. Puisque ces protéines sont structuralement apparentées aux protéines H2 de classe I, on peut considérer ces gènes comme une extension de la région du chromosome attachée aux fonctions concernées par le développement des lymphocytes et des macrophages.

Les antigènes de transplantation (molécules de classe I du CMH) sont des protéines transmembranaires constituées de deux chaînes dont la structure est illustrée dans la Figure 32.20.

* N. d. T. : On sait maintenant que les molécules d'histocompatibilité sont en fait des récepteurs pour l'antigène, le présentant au récepteur T. Les molécules de classe I présentent l'antigène aux lymphocytes cytotoxiques, alors que les molécules de classe II présentent l'antigène aux cellules auxiliaires de la réponse immunitaire, jouant notamment un rôle dans la coopération entre les lymphocytes T et B. L'antigène présenté par ces molécules semble être un peptide de 12 à 20 acides aminés, produit de digestion de l'antigène immunisant.

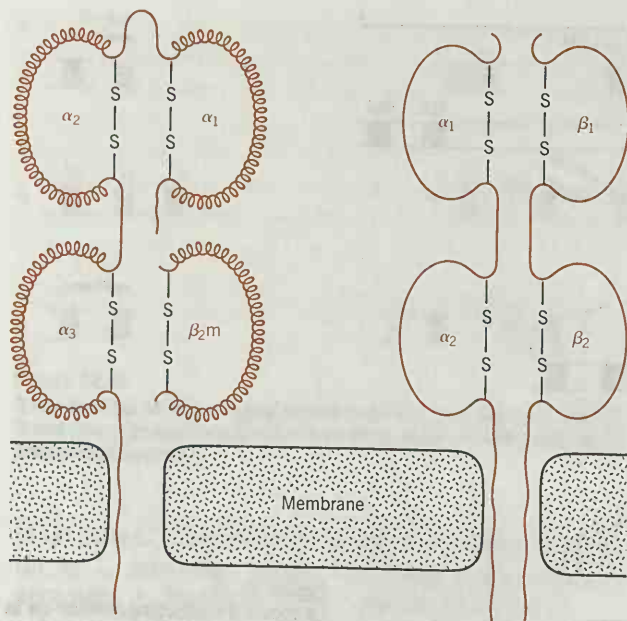


Figure 32.20

Les antigènes d'histocompatibilité de classe I et de classe II ont une structure similaire. Les antigènes de classe I sont constitués d'un seul polypeptide (α), avec trois domaines externes, qui interagit avec la microglobuline β_2 . Les antigènes de classe II sont constitués de deux polypeptides (α et β) dont l'interaction engendre une structure globale similaire.

Une des chaînes est la microglobuline β_2 de 12 000 daltons (codée par un seul gène localisé sur un autre chromosome). Ce composant est nécessaire pour que la protéine dimérique réside à la surface cellulaire.

Le composant transmembranaire est une chaîne de 45 000 daltons codée par le locus d'histocompatibilité. La protéine possède trois **domaines externes** (chacun d'une longueur d'environ 90 acides aminés, dont un interagit avec la microglobuline β_2), une **région transmembranaire** d'environ 40 résidus, et un court **domaine cytoplasmique** d'environ 30 résidus qui se trouve à l'intérieur de la cellule.

L'organisation des gènes de classe I coïncide avec la structure de la protéine. Le premier exon code pour une séquence signal (retirée de la protéine pendant le transfert membranaire). Les trois exons suivants codent pour chacun des domaines externes. Le cinquième exon code pour le domaine transmembranaire. Et les trois derniers petits exons codent ensemble pour le domaine cytoplasmique. Les gènes des antigènes de transplantation humains en diffèrent par leur domaine cytoplasmique qui est codé par deux exons uniquement.

Les gènes murins de classe I forment des groupes. Un total de 32 gènes occupant environ 850 kb d'ADN ont été jusqu'à maintenant identifiés (sous la forme de groupes séparés qui doivent encore être connectés). Le groupe le plus important a 7 gènes éparpillés sur 190 kb. Tous les gènes à l'intérieur de chaque groupe sont orientés dans la même direction ; des gènes adjacents tendent à être plus étroitement apparentés, ce qui suggère qu'ils proviennent de duplications ancestrales en tandem.

Les différences dans les schémas de restriction de l'ADN de différentes souches de souris peuvent être utilisées pour rattacher les localisations des différents groupes aux loci génétiques. (Ceci implique essentiellement l'utilisation des différences de restriction comme marqueurs dans la cartographie génétique). De cette façon, la plupart des gènes de classe I ont été cartographiés aux loci *Qa* et *TLa* ; seuls 5 des gènes se trouvent dans le groupe *H2*.

Ce résultat implique que les gènes de classe I sont présents aux deux extrémités de ce vaste groupe ; quant aux gènes de classe II et de classe III, soit ils se développent à l'intérieur du groupe, soit ils furent transloqués à leurs positions actuelles. La majorité du polymorphisme dans les différents gènes apparaît dans ceux du locus *H2*. La variation dans le nombre de gènes entre les différentes souches de souris semble apparaître largement dans les loci *Qa* et *TLa*.

Qu'est-ce qui est responsable du haut degré de polymorphisme dans ces gènes ? La plus grande variation de séquence entre allèles apparaît dans les premier et second domaines externes, prenant quelquefois la forme d'un groupe de substitutions de bases dans une petite région. Un des mécanismes impliqués dans leur apparition peut être une conversion génique entre gènes de classe I.

Des pseudogènes sont présents en même temps que les gènes fonctionnels ; à l'heure actuelle nous avons encore du chemin à faire avant d'estimer le nombre total de gènes actifs dans la région.

L'exon codant pour le troisième domaine externe des gènes de classe I est hautement conservé par rapport aux autres exons. Le domaine conservé représente probablement la région qui interagit avec la microglobuline β_2 , ce qui explique la nécessité d'une structure constante. Ce domaine présente également des homologues avec les domaines de la région constante des immunoglobulines.

Comme les gènes de la classe I, les gènes de la classe II et de la classe III sont également interrompus

avec les exons correspondants aux domaines protéiques. Il y a peu de gènes dans ces classes, environ 10 dans chacune. Les antigènes de la classe II sont constitués de deux chaînes, α et β , dont la combinaison produit une structure globale similaire à celle des protéines de la classe I.

Le gène de la microglobuline $\beta 2$ a été identifié. Il possède quatre exons, le premier codant pour une séquence signal, le second pour l'ensemble de la protéine (acides aminés 3 à 95), le troisième pour les quatre derniers acides aminés et une partie de la queue non traduite, et le dernier pour le reste de la queue.

La longueur du gène de la microglobuline $\beta 2$ est similaire à celle d'un gène d'immunoglobuline ; il y a certaines similitudes dans la constitution en acides aminés ; et il existe certaines homologues (limitées) de séquence nucléotidique entre la microglobuline $\beta 2$ et les domaines constants Ig ou les troisièmes domaines externes des gènes de type I. Il est possible que tous les groupes de gènes dont nous avons discuté dans ce

chapitre aient dérivé d'un ancêtre commun qui codait pour un domaine primitif.

POUR EN SAVOIR PLUS

Une vue générale des mécanismes impliqués dans l'apparition de la diversité des Ig a été donnée par **Tonegawa** (*Nature* **302**, 575-581, 1983). Un modèle de type délétion pour la jonction V-J a été présenté par **Max, Seidman & Leder** (*Proc. Nat. Acad. Sci. USA* **76**, 3340-3344, 1979) ; un modèle basé sur l'échange entre chromatides sœurs a été proposé par **Van Ness et al.** (*Proc. Nat. Acad. Sci. USA* **79**, 262-266, 1982). L'organisation des gènes Ig a été résumée par **Honjo & Habu** (*Ann. Rev. Biochem.* **54**, 803-830, 1985). Deux revues sur la commutation de classe ont été réalisées par **Davis, Kim & Hood** (*Cell* **22**, 1-2, 1980) et par **Shimizu & Honjo** (*Cell* **36**, 301-303, 1984). Une revue sur la mutation somatique a été réalisée par **Baltimore** (*Cell* **26**, 295-296, 1981). Les gènes des récepteurs des cellules T ont été analysés par **Hood, Kronenberg & Hunkapiller** (*Cell* **40**, 225-229, 1985). Une revue sur l'organisation des gènes d'histocompatibilité a été réalisée par **Steinmetz & Hood** (*Science* **222**, 727-732, 1983).

CHAPITRE 33

CHANGER L'ORGANISATION GÉNIQUE DU DEDANS ET DU DEHORS

Les réarrangements de l'ADN engendrent une diversité à la fois chez les eucaryotes et les procaryotes. Parfois le réarrangement crée de nouveaux gènes, nécessaires à une expression dans des circonstances particulières, comme dans le cas des immunoglobulines (voir Chapitre 32). Dans d'autres cas, les réarrangements sont responsables de la commutation de l'expression d'un gène préexistant en un autre ; ici le réarrangement est un mécanisme de régulation de l'expression génique.

La commutation entre les types sexuels chez la levure constitue un cas aux conséquences relativement simples. *S. cerevisiae* haploïde peut avoir l'un ou l'autre des deux types sexuels, et la transition de l'un vers l'autre est réalisée par une substitution physique d'ADN au locus qui détermine le type sexuel. Dans son résultat cette transition ressemble à celle de l'inversion de phase bactérienne, avec une simple commutation réalisée à partir du type existant vers une seule alternative. Le mécanisme, toutefois, implique une substitution de séquences.

Une situation où un réarrangement de l'ADN rend possible un certain nombre de variations, est fournie par les trypanosomes africains, parasites unicellulaires qui échappent à la réponse immunitaire de l'hôte en faisant varier leurs antigènes de surface. Il semble cohérent que le mécanisme utilisé pour combattre la flexibilité de l'appareil immunitaire soit analogue à celui utilisé pour

engendrer la diversité immunitaire : il repose sur un changement des séquences qui sont exprimées, par des réarrangements physiques dans le génome.

Finalement, nous avons des interactions parasites-symbiote-hôte dans lesquelles de l'ADN exogène provenant d'une bactérie est introduit dans une cellule végétale pour accroître sa capacité génétique.

LA LEVURE POSSÈDE DES LOCI MUETS ET DES LOCI ACTIFS POUR LE TYPE SEXUEL

La levure *S. cerevisiae* peut se propager avec succès aussi bien dans les conditions haploïdes que diploïdes. La conversion entre ces deux états se réalise par conjugaison (fusion des spores haploïdes pour donner un diploïde) et par sporulation (méiose des diploïdes pour donner des spores haploïdes). La capacité à s'engager dans ces activités est déterminée par le **type sexuel** de la souche.

Les propriétés des deux types sexuels sont résumées dans le **Tableau 33.1**. On peut considérer qu'elles s'appuient sur la proposition téléologique suivante : il ne sert à rien de s'accoupler à moins que les haploïdes soient de types génétiques différents ; et il ne sert à rien de sporuler à moins que le diploïde soit hétérozygote et qu'ainsi il puisse produire des recombinants.

Tableau 33.1
Le type sexuel contrôle plusieurs activités.

	$MATa$ ou $MATa/MATa$	$MAT\alpha$ ou $MAT\alpha/MAT\alpha$	$MATa/MAT\alpha$
Type cellulaire	a	α	a/ α
Croisement	Oui	Oui	Non
Sporulation	Non	Non	Oui
Sécrétion	Facteur a	Facteur α	Aucune

Les deux types sexuels sont déterminés par les allèles $MATa$ et $MAT\alpha$ du locus type sexuel. Les cellules haploïdes doivent être soit $MATa$ soit $MAT\alpha$. Les cellules de type opposé peuvent s'accoupler ; les cellules de même type ne peuvent pas.

La reconnaissance des cellules de type sexuel opposé est accomplie par la sécrétion de **phéromones**. Les cellules $MAT\alpha$ sécrètent le petit facteur polypeptidique α ; les cellules $MATa$ sécrètent un facteur a. Chaque

facteur agit sur les cellules de type sexuel opposé pour coordonner la fusion cellulaire et nucléaire.

Les diploïdes homozygotes ne peuvent pas sporuler. En revanche, les diploïdes de type $MATa/MAT\alpha$, produits par accouplement des haploïdes, peuvent sporuler. La **Figure 33.1** montre comment ce modèle permet de maintenir le cycle vital normal haploïde/diploïde.

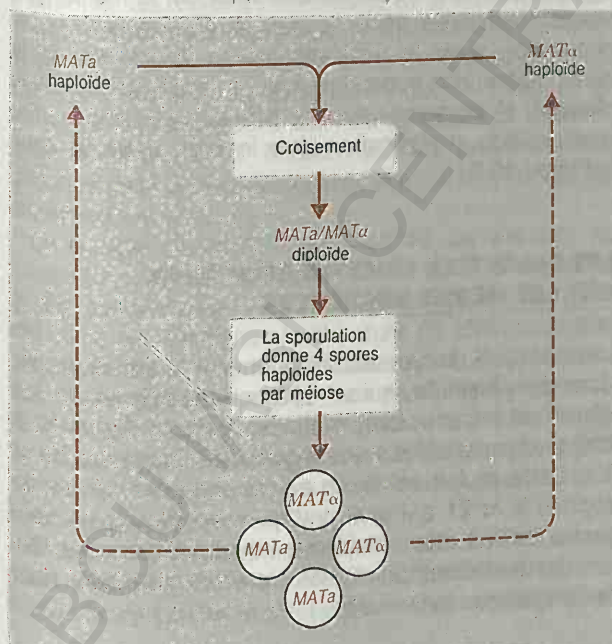
La **commutation** des types sexuels constitue un caractère singulier chez certaines souches de levures. Ces souches portent un allèle dominant HO et *changent leur type sexuel fréquemment*, une fois par génération. Les souches avec l'allèle récessif ho ont un type sexuel stable, dont la fréquence de changement est d'environ 10^{-6} .

L'expression du gène HO est elle-même sous le contrôle du type sexuel, puisqu'il ne fonctionne pas dans les diploïdes $MATa/MAT\alpha$ (téléologiquement, on peut penser qu'il n'y a pas besoin de commutation quand les deux allèles MAT sont de toute façon exprimés). Quand HO est présent dans une population de levures, quel que soit le type sexuel initial, en très peu de générations il y a un grand nombre de cellules des deux types sexuels, conduisant à la formation des diploïdes $MATa/MAT\alpha$ qui envahissent la population. La formation de diploïdes stables à partir d'une population haploïde peut être vue comme une *raison d'être* de la commutation.

L'existence d'une commutation suggère que *toutes les cellules contiennent l'information potentielle nécessaire pour être soit $MATa$ soit $MAT\alpha$, mais expriment un seul type*. Cette conclusion est confirmée par l'analyse de mutants pour chaque type d'allèle MAT . La commutation dans les souches HO permet à l'allèle muté d'être remplacé par l'un des gènes de type sauvage $MATa$ ou $MAT\alpha$.

D'où vient l'information de type sauvage ? Deux loci supplémentaires sont nécessaires pour la commutation. $HML\alpha$ est nécessaire pour que la commutation donne un type $MAT\alpha$; $HMRa$ est nécessaire pour que la commutation donne un type $MATa$. Ces loci se trouvent sur le

Figure 33.1
Le cycle vital de la levure évolue via la conjugaison des haploïdes $MATa$ et $MAT\alpha$ pour donner des hétérozygotes diploïdes qui sporulent pour engendrer des spores haploïdes.



même chromosome que *MAT*; *HML* est loin, à gauche, *HMR* loin, à droite.

Le **modèle à cassette** pour le type sexuel propose que *MAT* a une **cassette active** de type α ou de type *a*. *HML* et *HMR* ont des **cassettes muettes** qui sont physiquement substituées à la cassette *MAT* par le processus de commutation. Habituellement *HML* porte une cassette α , alors que *HMR* porte une cassette *a*. Les cassettes portent l'information codant pour le type sexuel.

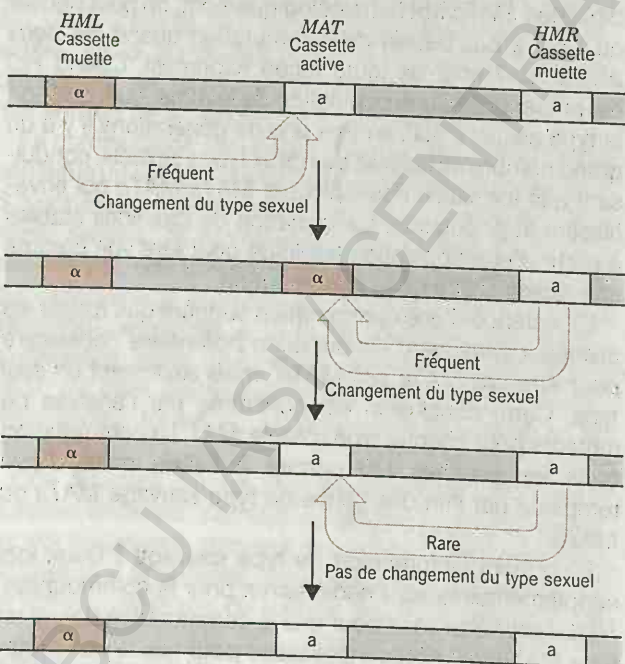
Le processus de commutation est illustré dans la **Figure 33.2**. La commutation est non réciproque, parce que la copie située au niveau *HML* ou *HMR* remplace l'allèle situé au niveau *MAT*. Nous le savons parce qu'une mutation au niveau de *MAT* est perdue définitivement quand il est remplacé par commutation, il ne s'échange pas avec la copie qui le remplace.

Les copies présentes au niveau *HML* ou *HMR* peuvent

Figure 33.2

Les changements du type sexuel se produisent quand les cassettes muettes remplacent les cassettes actives de génotype opposé ; quand des transpositions se produisent entre des cassettes de même type, le type sexuel reste inchangé.

(Des études avec des loci mutés montrent que *HML α* remplace parfois *MAT α* et que *HMR a* remplace parfois *MAT a* . De telles commutations passent inaperçues à moins qu'une mutation soit présente à un des loci, pour le marquer génétiquement.)



aussi être mutées. Dans ce cas la commutation introduit un allèle muté au locus *MAT*. La copie mutée au niveau *HML* ou *HMR* reste là à travers un nombre indéfini de commutations. Comme dans la transposition répliquative, l'élément donneur engendre une nouvelle copie au niveau du site receveur alors que lui-même reste inviolé.

La commutation du type sexuel est un événement dirigé, dans lequel il y a un seul receveur (*MAT*), mais deux donneurs potentiels (*HML* et *HMR*). La commutation change habituellement le type sexuel. Elle implique le remplacement de *MAT α* par la copie de *HML α* ou le remplacement de *MAT a* par la copie de *HMR a* . Dans 80 à 90 % des commutations, l'allèle *MAT* est remplacé par celui du type opposé, un effet qui est apparemment déterminé par le phénotype de la cellule. Les cellules de phénotype *a* choisissent préférentiellement *HML* comme donneur ; les cellules de phénotype α choisissent préférentiellement *HMR*.

(Il est possible d'obtenir des souches de levure chez lesquelles l'orientation usuelle des cassettes muettes est inversée. Quand leurs génotypes sont *HML α* et *HMR a* , 92 % des remplacements sont homologues : une cassette *a* est remplacée par une autre cassette *a* ou une α est remplacée par une autre α , parce que le choix des cassettes donneuses reste le même quel que soit le contenu des loci muets.)

Sa dépendance vis-à-vis de la lignée cellulaire est un des caractères intéressants de la commutation. La commutation est détectée uniquement dans les produits d'une division ; les deux cellules filles ont le même type sexuel, dérivé de celui du parent. Ceci suggère une connection avec la réplication dans la mesure où la commutation doit apparaître avant que le locus *MAT* soit répliqué. (A comparer avec la connexion existant entre transposition et réplication chez le maïs, montrée dans la Figure 30.6.)

LES CASSETTES MUETTES ET ACTIVES ONT LES MÊMES SÉQUENCES

Les deux loci du type sexuel muet et actif ont été clonés dans des plasmides qui se répliquent dans la levure. Les gènes ont les structures montrées dans la **Figure 33.3**. Chaque type d'allèle a une région distincte, notée *Y α* ou *Y a* . Cette région est flanquée de chaque côté par les régions *X* et *Z1* qui sont pratiquement identiques dans toutes les cassettes. (Seulement quatre différences, toutes des mutations ponctuelles, ont été identifiées dans les séquences individuelles.)

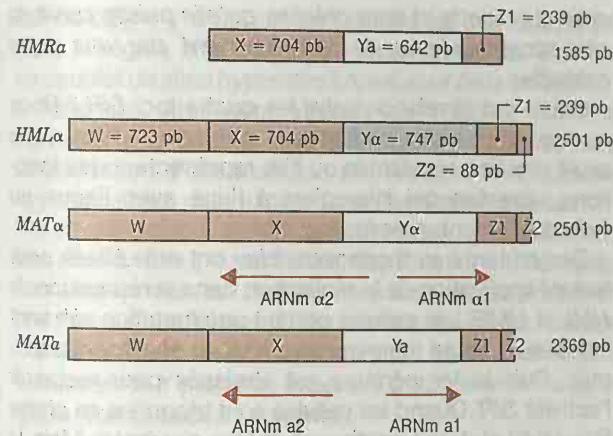


Figure 33.3
Les cassettes muettes ont les mêmes séquences que les cassettes actives correspondantes. Seule la région Y change entre les types a et α . HMR ne possède pas les séquences W et $Z2$ communes à HML et MAT .

Les mutations dans $MAT\alpha$ constituent deux groupes de complémentation ($\alpha1$ et $\alpha2$), alors que les mutations dans $MATa$ en constituent un seul groupe ($a1$). La capacité des cassettes portées par le plasmide à compléter avec les loci mutés au niveau MAT permet d'assimiler les régions individuelles de chaque locus aux groupes de complémentation (en introduisant une anomalie à des sites spécifiques dans les copies clonées). La relation entre les groupes de complémentation et les transcrits correspondant aux allèles MAT est indiquée dans la Figure 33.3. Notons que les protéines sont déduites de la séquence codante et n'ont pas encore été caractérisées.

Le groupe de complémentation $a1$ est assimilé au transcrit $a1$, qui est entièrement codé dans la région Ya et est donc unique pour le locus de type a . Le gène est donc unique pour le locus de type a . Le gène $MATa1$ a deux introns ; la structure du (des) produit(s) polypeptidique(s) est encore à déterminer. Aucune fonction n'a été détectée pour le transcrit $a2$, qui est transcrit sous le contrôle d'un promoteur dont au moins une partie se trouve dans Ya . Le transcrit se trouve dans la région commune X et pourrait être traduit sous forme de deux courts polypeptides.

Chacun des transcrits produits au site $MAT\alpha$ correspond à un des groupes de complémentation connus. Les transcrits sont initiés et s'allongent en sens opposé dans

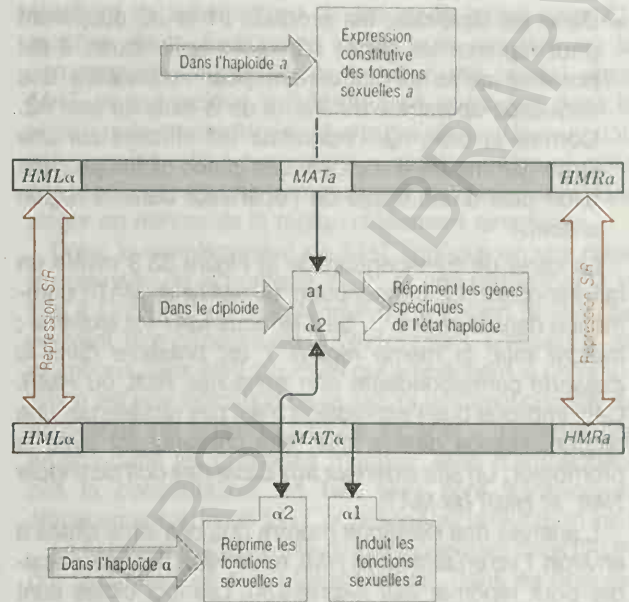


Figure 33.4
Dans les diploïdes, les fonctions $a1$ et $\alpha2$ coopèrent pour réprimer les fonctions spécifiques de l'état haploïde. Chez les haploïdes, la fonction $\alpha2$ réprime les fonctions sexuelles a ; $\alpha1$ induit les fonctions sexuelles α .

$Y\alpha$ et chacun s'étend dans la région commune adjacente (X et $Z1$). Le transcrit $\alpha1$ est traduit en une protéine largement codée par $Y\alpha$. Le transcrit $\alpha2$ peut être traduit à partir d'un codon d'initiation localisé dans $Y\alpha$, mais la majorité de la séquence correspond à X .

Les fonctions des produits, déduites des déficits des mutants présents dans chaque groupe de complémentation, sont illustrées dans la Figure 33.4. Elles régulent la transcription des autres gènes, et incluent les deux types de régulation, positive et négative :

- Chez les haploïdes a , les fonctions de conjugaison a sont exprimées constitutivement, mais les fonctions des produits de $MATa$ sont inconnues.
- Chez les haploïdes α , le produit $\alpha1$ active des gènes spécifiques α dont les produits sont nécessaires au type sexuel α . Le produit $\alpha2$ réprime les gènes responsables de la production du type sexuel a , en se liant à une séquence opératrice localisée en amont des gènes cibles. L'opérateur a la propriété inhabituelle de pouvoir fonctionner apparemment n'importe où dans la région située en amont du promoteur.

- Dans les diploïdes, les produits $\alpha 1$ et $\alpha 2$ coopèrent pour réprimer les gènes haploïdes spécifiques. Il est possible qu'ils se combinent pour reconnaître une séquence opératrice différente de la cible du seul $\alpha 2$. Comme la cible $\alpha 2$, l'opérateur est efficace sur une large gamme de distances. Les gènes cibles peuvent avoir plus d'une copie de l'opérateur dans la région amont.

La carte de transcription de la Figure 33.3 révèle un fait intrigant. La transcription de $MATa$ et $MAT\alpha$ commence dans la région Y. Seul le locus MAT est exprimé ; *malgré tout, la même région Y est présente dans la cassette correspondante non transcrite, HML ou HMR.* Ceci implique que l'expression n'est pas régulée par une reconnaissance directe d'un site chevauchant avec le promoteur. *Un site extérieur aux cassettes doit distinguer HML et HMR de MAT.*

L'analyse des délétions montre que des sites situés à environ 1 kb en amont de HML et de HMR sont nécessaires pour réprimer leur expression. Les loci cibles sont quelquefois appelés E_L (près de HML) et E_R (près de HMR). Ces sites de contrôle ont deux propriétés intrigantes :

- Ils se comportent comme des enhanceurs négatifs *, dans le sens où ils peuvent fonctionner à distance (jusqu'à 2,5 kb d'un promoteur) et dans chaque direction.
- Ils sont associés avec les séquences *ars*, qui fonctionnent probablement comme origines de réplication.

Peut-on trouver la base du contrôle de l'activité des cassettes en identifiant d'autres loci qui agiraient sur les cassettes ? Certains loci sont révélés par des mutations qui empêchent l'expression de l'allèle occupant le site MAT . Des mutations dans d'autres loci permettent aux cassettes habituellement muettes situées au niveau HML et HMR d'être exprimées. (Quand ceci aboutit à l'expression des deux fonctions a et α , les cellules se comportent comme des diploïdes $MATa/MAT\alpha$.)

Quatre groupes de complémentation ont été identifiés dans lesquels les mutations conduisent à l'expression de HML et HMR . Elles sont appelées *SIR* (régulateur d'information muet = silent information regulator). Les quatre loci *SIR* de type sauvage sont nécessaires pour maintenir HML et HMR dans leur état réprimé ; une mutation dans n'importe lequel de ces loci qui donne un allèle *sir*⁻ a deux effets. HML et HMR peuvent être tous deux transcrits. Et les deux cassettes muettes deviennent des cibles pour être remplacées par commutation. *Ainsi le même événement régulateur est impliqué pour réprimer une*

cassette muette et pour prévenir qu'elle puisse servir de site accepteur pour le remplacement par une autre cassette.

Quelle est la relation entre les quatre loci *SIR* ? Nous ne savons pas s'ils codent pour des sous-unités d'une seule protéine régulatrice ou s'ils représentent des fonctions séparées qui interagissent l'une avec l'autre ou individuellement avec les loci cibles.

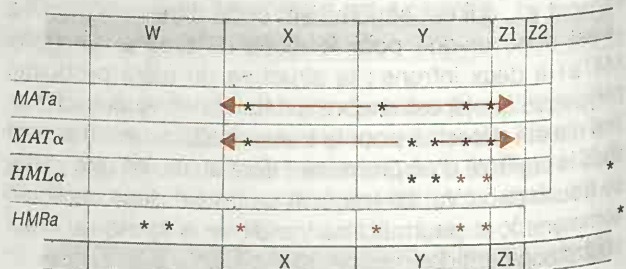
Des mutants *sir* thermosensibles ont été utilisés pour tester l'implication de la réplication dans la répression de HML et HMR . Les cellules portant une mutation *sir*^{ts} sont incubées à haute température ; HML et HMR sont exprimés. Puis la température est abaissée pour restaurer l'activité *SIR*. Quand les cellules sont bloquées en phase G1, HML et HMR continuent à être exprimés. Mais la répression peut être établie dès que la réplication de l'ADN est autorisée. Ceci suggère que *la réplication des séquences ars associées aux éléments E est nécessaire pour établir la répression de HML et HMR.* Ceci est une nouvelle forme de répression, et nous ne savons pas encore comment cela fonctionne au niveau moléculaire.

L'activation d'une cassette conduit à des changements dans la structure de sa chromatine. Toutes les cassettes possèdent des sites hypersensibles à l'ADNase I, mais il y a des différences significatives entre les cassettes actives et muettes. De même, il y a des changements dans l'échelonnement des nucléosomes et dans le degré de superenroulement, qui peuvent être associés à l'activation des gènes.

Les sites hypersensibles perturbés sont résumés dans la Figure 33.5. Le schéma général de l'hypersensibilité est similaire dans $MATa$ et $MAT\alpha$. Dans chaque cas, il y a un site à l'extrémité 3' du transcrit d'orientation gauche

Figure 33.5

La majorité des sites hypersensibles à l'ADNase I trouvés dans MAT ne sont pas présents dans HML ou HMR ; mais ils sont activés dans les mutants *sir*⁻. Certains sites hypersensibles permanents sont trouvés dans les régions bordantes HML et HMR .



* Les sites hypersensibles sont toujours présents

* Les sites hypersensibles sont présents uniquement dans les souches *sir*⁻

* N. d. T. : Le nom de « silencer » a été donné à ce type d'élément d'ADN.

(a2 ou $\alpha 2$). Il y a un site soit simple (a) soit double (α) dans la région située entre les deux transcrits. Et il y a un doublet de sites hypersensibles autour de la limite Y-Z.

Tous ces sites sont absents de *HML* et *HMR*, à l'exception du doublet situé entre $\alpha 1$ et $\alpha 2$. Cependant, l'introduction d'une mutation *sir*⁻ restaure tous les sites hypersensibles manquants. Ainsi l'activation des cassettes est associée à la présence des sites hypersensibles. Les divers sites peuvent avoir des fonctions différentes. Les sites à l'intérieur de Y peuvent être nécessaires (bien que non suffisants) pour initier la transcription. La fonction des sites situés dans X est inconnue. Le doublet situé à la frontière X-Y peut être important pour initier la transposition (voir la section suivante).

Des sites hypersensibles sont trouvés dans les régions E près de *HML* et *HMR*. Leur présence ne dépend pas de la fonction *SIR* ; en fait, ces sites pourraient identifier les cassettes muettes comme cibles pour l'action de *SIR*. Quand ces sites sont reconnus par *SIR*, un signal doit être transmis pour éviter la formation de sites hypersensibles à l'intérieur des cassettes.

Comment les fonctions *SIR* agissent-elles à distance ? La transmission d'un effet structural le long de l'ADN est une possibilité. Par exemple, le degré de superenroulement pourrait influencer des sites dans les cassettes, ou un processus de mise en phase des nucléosomes pourrait déterminer si un site est disponible.

LA TRANSPPOSITION UNIDIRECTIONNELLE EST INITIÉE PAR LE LOCUS RÉCEPTEUR *MAT*

Dans les populations subissant la commutation, la recombinaison entre *MAT* et *HML* ou *HMR* se produit avec une faible fréquence. L'apparition de cet événement (aberrant) suggère que la commutation implique l'appariement entre *MAT* et l'une des cassettes muettes. Le résultat habituel de leur interaction est une commutation du type sexuel, mais, occasionnellement, une recombinaison réciproque apparaît à la place. Au premier chef la recombinaison réunit les loci recombinants et excise le matériel situé entre ceux-ci ; ceci signifie que les trois cassettes sont dans la même orientation sur le chromosome.

Les sites nécessaires à la transposition ont été identifiés grâce à des mutations qui empêchent la commutation. Des délétions à l'extrémité droite de Y n'ont aucun effet tant qu'elles ne traversent pas la frontière avec Z1 ; après elles empêchent la commutation. Les mutations ponctuelles actives en *cis* appelées *inc* (pour inconvertible) ont été trouvées au niveau de *MATa* et de *MATa*. Elles

résultent de substitutions de paires de bases dans Z1, proches de la limite avec Y. Dans le cas rare où la commutation se produit, la mutation est perdue dans *MAT*, ce qui indique qu'elle pourrait se trouver dans la région qui est remplacée. Une autre classe de mutation est appelée *stk* (pour stuck = collé) ; celles-ci sont conservées après une commutation, ainsi elles semblent siéger en dehors de la région réellement remplacée.

Donc le remplacement de *MAT* implique un site dans le voisinage de la frontière Y-Z1, et des séquences à la fois à l'intérieur et à l'extérieur de la région remplacée peuvent être nécessaires à l'événement. L'absence de mutations dans *HML* ou *HMR* qui empêchent la commutation est une indication de la nature unidirectionnelle du processus ; et si des mutations *inc* sont transférées par artifice dans les cassettes muettes, elles n'empêchent pas la commutation. En fait, elles sont transposées efficacement au locus *MAT*, où réapparaît la mutation *inc* !

La frontière Y-Z est le site d'un changement dans l'ADN qui peut être relié à une transposition. Dans les populations de cellules qui subissent la commutation, 1 à 3 % de l'ADN du locus *MAT* a une coupure du double brin à ce site. La coupure se trouve à 10 pb de la frontière et coïncide avec le site hypersensible à l'ADNase.

Nous pouvons maintenant suggérer une série possible d'événements expliquant la commutation. Le site hypersensible situé à la frontière Y-Z1 de *MAT* peut être accessible parce qu'il est libre de nucléosome. Il est reconnu par une endonucléase codée par le locus *HO*. En fait, ceci équivaut à suggérer que l'hypersensibilité à l'ADNase I *in vitro* reflète une sensibilité naturelle à l'endonucléase *HO in vivo* !

L'endonucléase *HO* — également appelée endo YZ — réalise une coupure décalée dans la région Z1, à 3 pb de la jonction YZ. Elle ne coupe pas les loci *MAT* mutés qui ne peuvent pas commuter. Les délétions qui laissent moins de 11 pb de la séquence cible ne sont pas clivées par l'enzyme, par conséquent le site de reconnaissance est relativement large pour une nucléase, environ 15 pb. Il est probable qu'il n'apparaisse qu'au niveau de *MAT* et au niveau de *HML* et *HMR* (où il est protégé de la coupure par les fonctions *SIR*).

La synthèse de l'endonucléase *HO* est sous contrôle du cycle cellulaire. La programmation de sa production peut expliquer la transmission caractéristique du type sexuel. Le gène est exprimé uniquement pendant la phase G1 d'une cellule mère, ce qui explique pourquoi les deux descendants ont le nouveau type sexuel.

La réaction déclenchée par le clivage est illustrée schématiquement en termes de simples brins dans la Figure 33.6. (Bien sûr, ceci doit être plus complexe en

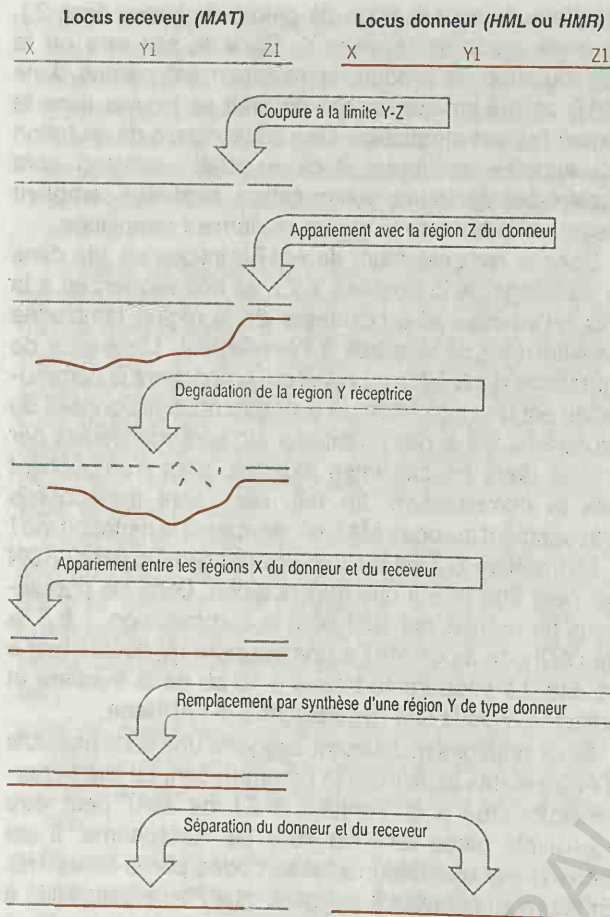


Figure 33.6

La substitution des cassettes est initiée par le locus receveur et peut impliquer l'appariement de chaque côté de la région Y avec le locus donneur. La séquence qui est retirée et remplacée peut s'étendre à la région X.

La figure montre la réaction en termes de simple brin tant pour le receveur que le donneur ; en réalité elle concerne les régions double-brins de l'ADN.

termes de doubles brins, et plusieurs voies sont possibles.) Supposons que l'extrémité libre Z de MAT envahisse le locus HML ou HMR et s'apparie avec la région Z. La région Y de MAT est dégradée jusqu'à ce qu'une région présentant une homologie avec X soit exposée. A ce point MAT est apparié avec HML ou HMR à la fois du côté gauche (X) et du côté droit (Z1). La région Y de HML ou de HMR est copiée pour remplacer la région Y de MAT (qui pourrait s'étendre au-delà des limites de Y lui-même). Les loci appariés se séparent. (L'ordre des événements pourrait être différent.) Le modèle est apparenté au schéma décrit dans la Figure 28.4 pour la recombinaison générale, et les étapes suivant la coupure

initiale requièrent les enzymes impliqués dans la recombinaison générale.

Comme dans le modèle de recombinaison par cassure du double brin, le processus est initié par MAT, le locus qui doit être remplacé. Dans ce sens, la description de HML et HMR comme loci donneurs se rapporte à leur rôle final mais pas au mécanisme du processus. Comme dans la transposition répllicative, le site donneur n'est pas affecté, mais un changement de séquence se produit au niveau du récepteur ; contrairement à la transposition, le locus receveur subit une substitution plutôt qu'une addition du matériel.

LES TRYPANOSOMES RÉARRANGENT L'ADN POUR EXPRIMER L'ANTIGÈNE DE SURFACE

Chez les êtres humains, la maladie du sommeil (et une maladie similaire chez les bovins) est causée par infection par des trypanosomes africains. Le parasite unicellulaire suit le cycle vital illustré dans la Figure 33.7, dans lequel il alterne entre la mouche tsé-tsé et le mammifère.

Au cours de ce cycle vital, le parasite subit plusieurs changements morphologiques et biochimiques. Le changement biochimique le plus significatif réside au niveau de la **glycoprotéine de surface variable (VSG = variable surface glycoprotein)**, le principal composant du manteau superficiel. Le manteau recouvre la membrane plasmique et est constitué d'une monocouche de $5 - 10 \times 10^6$ molécules d'une seule VSG, qui est la seule structure antigénique exposée à la surface. Un trypanosome exprime une seule VSG à la fois, et sa capacité à changer de VSG lui permet de survivre au cours du cycle infectieux mouche-mammifère.

Considérons le cycle au début, quand une mouche récupère un trypanosome en piquant un mammifère infecté. Le trypanosome ingéré perd sa VSG. Après trois semaines environ, sa descendance se différencie en « forme métacyclique », qui retrouve un manteau VSG. Cette forme est injectée par la mouche dans le flux sanguin d'un mammifère au cours d'une piqûre. Le trypanosome se multiplie dans le sang du mammifère. Sa descendance continue d'exprimer la VSG métacyclique pendant environ une semaine. Ensuite une nouvelle VSG est synthétisée, et d'autres transitions se produisent toutes les 1 à 2 semaines.

Chaque VSG est immunologiquement distincte. Le processus de transition est appelé **variation antigénique** ; il en résulte que l'antigène, qui doit être reconnu

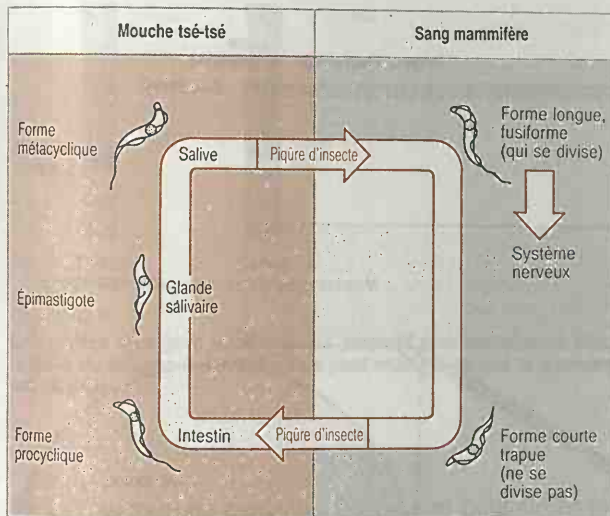


Figure 33.7
T. brucei passe par différentes formes morphologiques quand son cycle vital alterne entre une mouche tsé-tsé et le mammifère hôte.

par le système immunitaire, change constamment. De cette façon, le trypanosome échappe à la surveillance immunitaire, et se perpétue indéfiniment. Chaque transition de la VSG est accompagnée par une nouvelle vague de parasitémie, avec des symptômes de fièvre, d'éruption, etc. ; les parasites envahissent éventuellement le système nerveux, après quoi le mammifère hôte devient progressivement plus léthargique et éventuellement comateux.

Les trypanosomes diffèrent quant à leur spécificité d'hôtes ; et l'espèce la mieux étudiée est une variété de *Trypanosoma brucei* qui se développe bien chez les animaux de laboratoire mais pas chez l'homme. Les souches de laboratoire de *T. brucei* commutent les VSG spontanément avec un taux de 10^{-4} - 10^{-6} par division. La commutation se produit indépendamment du système immunitaire de l'hôte, mais de nouveaux variants sont sélectionnés par le système, parce qu'il est incapable de reconnaître et d'agir contre la nouvelle VSG.

Quelle est la structure de la VSG ? Les séquences des protéines (et de l'ADNc) donnent la vue générale de la structure de la VSG montrée dans la **Figure 33.8**. Une VSG native est longue d'environ 500 acides aminés ; elle possède une séquence signal, suivie d'une longue région variable, et enfin suivie par un des trois types de région d'homologie qui se termine par une courte queue hydrophobe. La séquence signal est clivée au cours de la sécrétion ; la queue hydrophobe est retirée avant que la

VSG atteigne la surface externe. Le COOH terminal de la VSG mature est attaché de façon covalente à la membrane du trypanosome, et les trois types de région d'homologie sont distingués selon leur acide aminé C-terminal. La région variable N-terminale fournit l'unique déterminant antigénique.

La structure exacte grâce à laquelle la VSG s'attache à la membrane n'est pas claire. Cependant, la VSG peut être libérée de la membrane par une enzyme qui enlève un acide gras ; cette réaction (qui est utilisée pour purifier la VSG) peut être importante *in vivo* en permettant le remplacement d'une VSG par une autre à la surface du trypanosome*.

Combien de variétés de VSG peuvent être exprimées par un trypanosome ? Il n'est pas clair qu'une limite quelconque soit atteinte avant la mort de l'hôte. Un trypanosome isolé peut fabriquer au moins 100 VSG tellement différentes quant à leur séquence que les anticorps dirigés contre l'une ne peuvent pas réagir contre les autres.

La variation des VSG est codée dans le génome du trypanosome. Chaque trypanosome porte le répertoire entier de VSG de sa souche. La diversité dépend donc du changement de l'expression d'un gène préexistant en un autre.

Le génome du trypanosome a une organisation inhabituelle, constituée d'un très grand nombre d'unités qui ségrègent. En plus d'un nombre inconnu de chromosomes, il contient environ 100 « mini-chromosomes » qui contiennent chacun quelque 50-150 kb d'ADN. Des expériences d'hybridation identifient approximativement 1 000 gènes VSG, dispersés parmi toutes les classes de taille du matériel chromosomique.

Chaque VSG est codée par un **gène dont il existe une copie d'origine**. Ces gènes peuvent être divisés en deux classes selon leur localisation chromosomique :

- Les gènes télomériques se trouvent à 5-15 kb d'un télomère. Il pourrait exister environ 200 de ces gènes si chaque télomère en possède un.
- Les gènes internes résident dans les chromosomes (plus formellement, ils sont à quelque 50 kb d'un télomère).

Comme on peut s'y attendre avec une grande famille de gènes, les copies d'origine individuelles montrent des degrés variables de relation, qui reflètent probablement leur origine par duplication et variation. Les gènes qui sont étroitement apparentés, et qui peuvent en fait déclencher la même réponse antigénique, sont appelés isogènes.

* N. D. T. : L'extrémité C-terminale des VSG semble attachée à la membrane par un pont phospho-inositol.

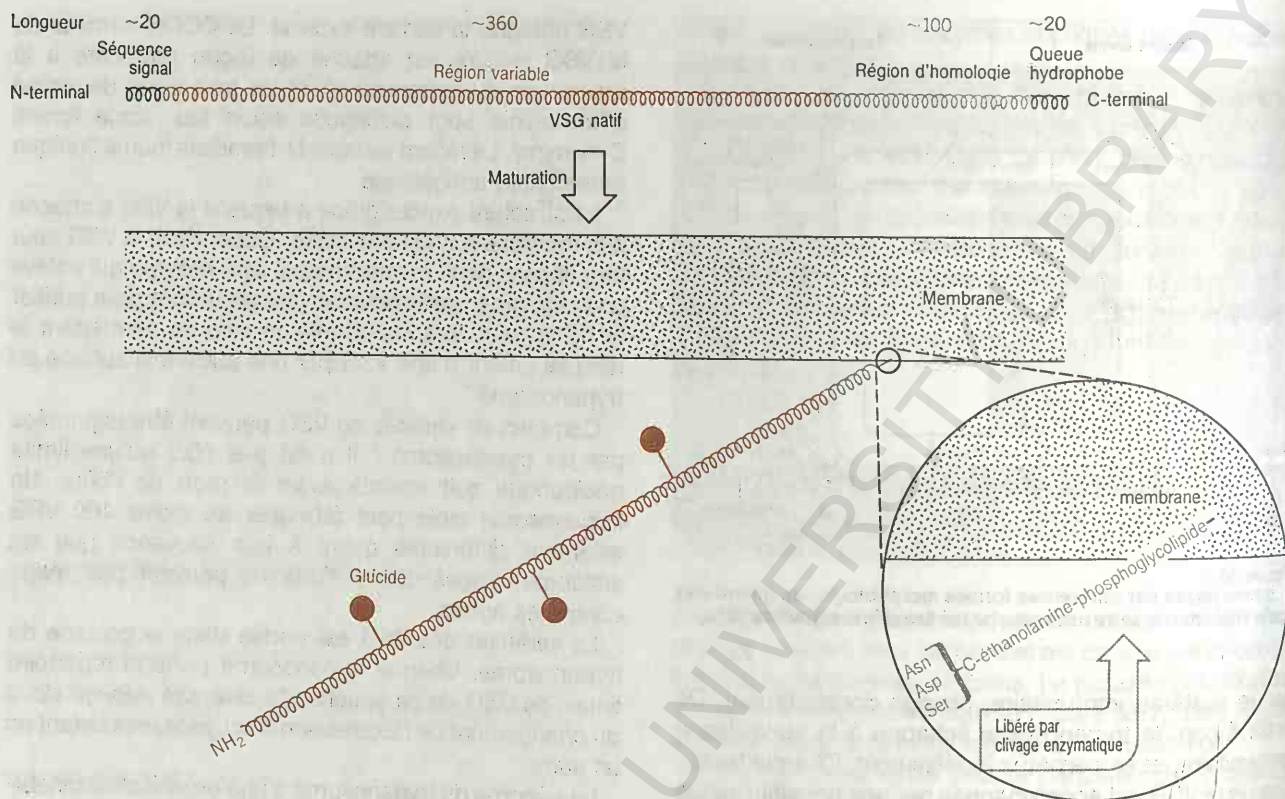


Figure 33.8

Chaque VSG est synthétisée sous forme de protéine native dont la séquence signal N-terminale et la queue hydrophobe C-terminale sont coupées pendant la maturation. L'extrémité C-terminale de la protéine mature est liée de façon covalente à la membrane par l'intermédiaire d'un glycolipide. Les résidus glucidiques sont attachés à l'intérieur de la molécule.

Comment un gène VSG particulier est-il sélectionné pour être exprimé ? Un seul gène VSG est transcrit dans un trypanosome donné. Un gène VSG exprimé a une propriété caractéristique : *il est situé près d'un télomère*. Ceci suggère immédiatement que la sélection d'un gène à exprimer va dépendre du fait que la copie de base est elle-même télomérique ou interne.

On peut suivre le devenir des gènes impliqués dans l'activation grâce à la carte de restriction. Une sonde représentant une séquence exprimée peut être dérivée de l'ARNm. Alors on peut déterminer le statut des gènes correspondant à la sonde. On peut observer des résultats différents pour les gènes de base télomériques et internes.

La Figure 33.9 montre que, quand un gène télomérique est activé, il n'est pas nécessaire que le nombre de gènes change. La structure du gène peut être essentiellement

non affectée comme on peut le détecter avec la carte de restriction. La taille du fragment contenant le gène peut varier légèrement, parce que la longueur du télomère change constamment.

La Figure 33.10 montre que quand un gène interne est activé, un nouveau fragment est trouvé. La copie d'origine du gène reste inchangée ; le nouveau fragment représente une copie du gène formée près du télomère. La copie supplémentaire est appelée **copie d'expression** (ELC = expression-linked copy : copie liée à l'expression) ; elle apparaît quand le gène est exprimé et disparaît quand le gène est éteint.

La formation de ELC se produit par conversion génique. Comme la commutation du type sexuel de la levure, elle représente le remplacement d'une « cassette » au locus actif (télomérique) par une cassette mise en réserve. Bien sûr, le système VSG est plus souple dans la mesure où

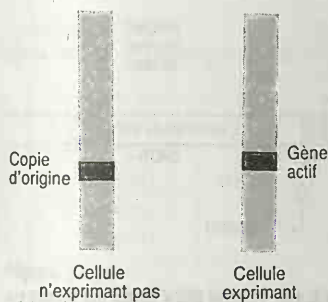


Figure 33.9

Les copies d'origine télomériques peuvent être activées *in situ* ; la taille du fragment de restriction peut changer quand le télomère est allongé.

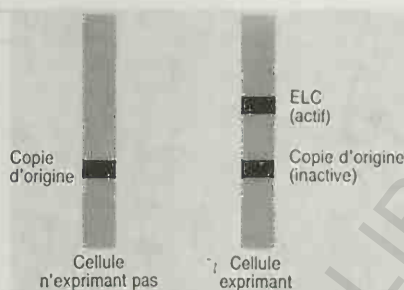


Figure 33.10

Les copies d'origine internes peuvent être activées seulement en provoquant une duplication du gène à un site lié à l'expression.

il y a beaucoup de cassettes donneuses potentielles (et aussi plus d'un seul récepteur potentiel).

Les copies d'origine internes ne peuvent être activées que par une seule voie : la duplication en une ELC. Les copies d'origine télomériques peuvent être activées au choix selon deux voies. Si elles utilisent la voie de la duplication, la copie d'origine reste à son télomère, alors qu'une copie liée à l'expression apparaît à un autre télomère (créant un nouveau fragment comme il est illustré pour les copies d'origine internes dans la Figure 33.10). Ou elles peuvent être activées *in situ*, auquel cas aucun changement n'apparaît dans le nombre de gènes (comme il est illustré dans la Figure 33.9).

Le seul gène VSG actif dans un trypanosome est localisé à un **site d'expression**. Les deux types d'événements qui peuvent créer un gène VSG actif sont illustrés dans la **Figure 33.11**.

- La duplication transfère la séquence d'une copie d'origine pour remplacer la séquence occupant jusqu'alors le site d'expression. Les copies internes ou télomériques peuvent être activées directement par duplication dans le site d'expression. *La substitution d'une cassette par une autre n'interfère pas avec l'activité du site.*
- L'activation *in situ* est possible uniquement pour une séquence déjà présente à un télomère, pour créer un nouveau site d'expression. *Quand un site télomérique est activé *in situ*, le site d'expression précédant doit cesser d'être actif et le nouveau site devient alors le site d'expression.*

La cible pour la transposition des copies d'origine internes n'est probablement pas limitée au site d'expression. Le même mécanisme peut être utilisé pour former

une copie télomérique du gène à un site non couramment exprimé.

On suppose qu'un changement se produit à un site d'expression quand il devient actif. Le changement pourrait être une altération dans la structure de la chromatine, ou pourrait impliquer un changement directement au niveau de l'ADN, peut-être une modification telle que la méthylation, ou peut-être une réorganisation telle que l'introduction d'une séquence activatrice (enhancer).

Combien existe-t-il de sites d'expression ? On sait que dans certains cas une série de copies d'origine peuvent être activées en remplaçant l'une par l'autre au même site d'expression. Dans d'autres cas (spécialement quand l'activation *in situ* des copies télomériques est en jeu), le site d'expression peut changer. Jusqu'à maintenant seuls quelques sites d'expression ont été observés, ce qui suggère que seul un sous-groupe de télomères peut fonctionner comme tel, mais il est possible qu'en fait n'importe quel télomère puisse être utilisé (bien que certains puissent être préférés à d'autres).

La structure du gène VSG au niveau de ELC est inhabituelle, comme l'illustre la **Figure 33.12**. La longueur de l'ADN transféré dans ELC est de 2 500 à 3 500 pb, un peu plus longue que la région codant pour VSG de 1 500 pb. La majorité de la longueur supplémentaire est en amont du gène. Les points de cross-over où la séquence dupliquée s'attache à ELC ne semblent pas être précisément déterminés. Au moins dans certains cas, les signaux pour l'initiation et la terminaison de la transcription (et quelquefois aussi la fin de la région codante elle-même) peuvent être fournis par les séquences bordantes.

L'analyse des événements au niveau de l'extrémité 5' de l'ARNm VSG est compliquée par le fait que l'ARN

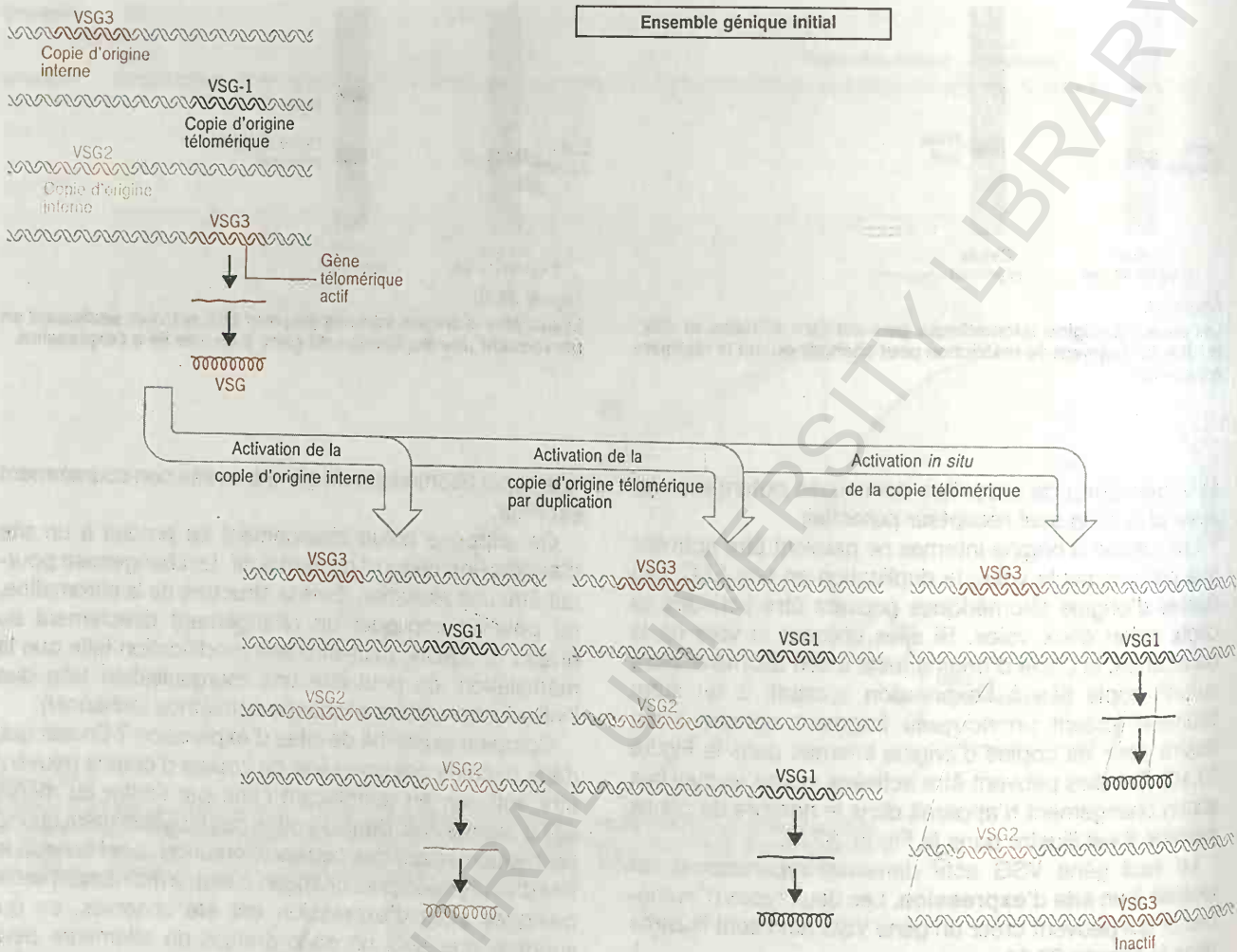


Figure 33.11

Les gènes VSG actifs peuvent être formés par transfert duplicatif d'une copie d'origine interne ou télomérique dans un site d'expression, ou par activation d'une copie télomérique déjà présente à un site d'expression potentiel.

mature commence avec une séquence de 35 bases codée ailleurs, et probablement ajoutée en *trans* à l'extrémité 5' nouvellement synthétisée (voir Chapitre 23). Cependant, un modèle suppose que la séquence insérée est activée du fait de son insertion en aval d'un promoteur actif*.

De chaque côté de la région transposée se trouve une vaste région qui n'est pas coupée par les enzymes de restriction. Ces « régions stériles » peuvent être constituées d'ADN répétitif; elles couvrent quelque 8 kb en

aval et jusqu'à 40 kb en amont de ELC. En aval, la région stérile est largement constituée de répétitions de la séquence CCCTAA, et va jusqu'au télomère. En amont, elle peut également être constituée de séquences répétitives, mais leur nature n'est pas encore claire. Cependant, l'existence des régions stériles a été un obstacle à la caractérisation des gènes ELC par clonage.

L'ordre dans lequel les gènes VSG sont exprimés au cours d'une infection est irrégulier. Ceci peut être un caractère important pour la survie du trypanosome. Si

* N. d. T. : L'hypothèse d'une réaction de trans-épissage semble maintenant largement admise.

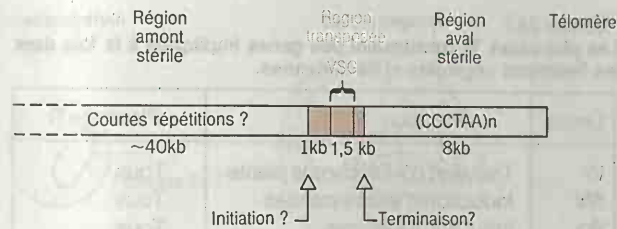


Figure 33.12

La copie du gène VSG liée au site d'expression contient des régions stériles de part et d'autre de la région transposée, qui s'étend d'environ 1 000 pb en amont de la région codant pour VSG jusqu'à un site localisé près de l'extrémité 3' de l'ARNm.

les gènes VSG étaient utilisés dans un ordre prédéterminé, un hôte pourrait repousser l'infection en développant une réaction contre l'un des éléments précoces. La nécessité d'imprévisibilité dans la production de VSG peut être responsable de l'évolution d'un système comportant de nombreuses séquences donneuses et de multiples récepteurs.

La variation antigénique n'est pas un phénomène spécifique des trypanosomes. La bactérie *Borrelia hermsii* provoque une fièvre récidivante chez l'homme et des maladies analogues chez d'autres mammifères. Le nom de la maladie reflète son évolution irrégulière : des périodes de troubles alternent avec des périodes de calme. Quand les fièvres apparaissent, des spirochètes sont trouvés dans le sang ; ils disparaissent pendant les périodes de calme, au moment où l'hôte répond avec des anticorps spécifiques.

Comme les trypanosomes, *Borrelia* survit en modifiant une protéine de surface, appelée protéine variable principale (VMP = variable major protein). Les modifications dans la VMP sont associées à des réarrangements dans le génome. On ne connaît pas encore l'étendue des variants codés ou les mécanismes utilisés pour altérer leur expression. Il est intrigant, toutefois, que l'eucaryote *Trypanosoma* et le procaryote *Borrelia* doivent tous les deux utiliser la variation antigénique comme moyen d'échapper à la surveillance immunitaire.

L'INTERACTION DE L'ADN DU PLASMIDE TI AVEC LE GÉNOME DE LA PLANTE

La galle du collet (Crown gall disease) peut être induite dans la plupart des plantes dicotylédones par la bactérie du sol *Agrobacterium tumefaciens*. La bactérie est un parasite qui effectue un changement génétique dans la cellule hôte eucaryote, avec des conséquences à la fois



Figure 33.13

Un *Agrobacterium* porteur d'un plasmide TI du type nopaline induit un tétatome, dans lequel se développent des structures différenciées.

Photographie aimablement fournie par Jeff Schell.

pour le parasite et l'hôte. Il améliore les conditions de survie du parasite. Et la cellule végétale se développe comme une tumeur. La Figure 33.13 est une photographie d'une telle tumeur.

Les *Agrobacteria* sont requises pour induire la formation de la tumeur, mais les cellules tumorales n'ont pas besoin de la présence continue des bactéries. Comme pour les tumeurs animales, les cellules végétales ont été transformées dans un état où de nouveaux mécanismes gouvernent la croissance et la différenciation.

Les tumeurs synthétisent de nouveaux dérivés de l'arginine appelés des opines, qui peuvent être utilisés comme la seule source carbonée et/ou azotée par la souche *Agrobacterium* inductrice. Les souches d'*Agro-*

bacterium peuvent être classées selon le(s) type(s) d'opines qui sont synthétisées dans les tumeurs qu'elles induisent.

Le principe inducteur de la tumeur réside dans le **plasmide Ti** présent chez *Agrobacterium*. Le plasmide porte les gènes qui contrôlent le catabolisme des opines chez *Agrobacterium* et les gènes qui déterminent la synthèse des opines dans les cellules végétales transformées. Les plasmides Ti (et donc les *Agrobacteria* dans lesquelles ils résident) peuvent être divisés en quatre groupes selon les types d'opines synthétisées :

Les **plasmides nopaline** portent les gènes contrôlant la synthèse de la nopaline dans les tumeurs et son utilisation dans les bactéries. Les tumeurs nopaline peuvent se différencier en excroissances avec des structures anormales. Elles ont été appelées **tératomes** par analogie avec certaines tumeurs de mammifères qui gardent la capacité de se différencier en structures embryonnaires précoces.

Les **plasmides octopine** sont identiques aux plasmides nopaline, mais l'opine correspondante est différente. Cependant, les tumeurs octopine sont habituellement non différenciées et ne forment pas de tératomes.

Les **plasmides agropine** portent les gènes contrôlant le métabolisme de l'agropine ; les tumeurs ne se différencient pas, se développent difficilement, et meurent rapidement.

Les **plasmides Ri** peuvent induire une maladie avec racines chevelues sur certaines plantes et une galle du collet sur d'autres. Ils ont des gènes de type agropine, et peuvent avoir des segments dérivés des deux plasmides nopaline et octopine.

Outre les gènes contrôlant la synthèse et l'utilisation des opines, un plasmide Ti porte des gènes nécessaires à diverses fonctions des bactéries et des cellules tumorales. Elles sont résumées dans le **Tableau 33.2**. Les gènes utilisés dans les bactéries codent pour la réplication du plasmide et l'incompatibilité, le transfert entre bactéries, la sensibilité aux phages, et la synthèse d'autres composés dont certains sont toxiques pour d'autres bactéries du sol. Les gènes utilisés dans la cellule végétale codent pour le transfert de l'ADN dans la plante, l'induction de l'état transformé, de l'excroissance et des racines.

La spécificité des gènes opine dépend du type du plasmide. Les gènes nécessaires pour la synthèse des opines sont liés aux gènes dont les produits catabolisent la même opine ; ainsi chaque souche d'*Agrobacterium* amène les cellules tumorales de la galle du collet à synthétiser des opines qui sont utiles à la survie du

Tableau 33.2

Les plasmides Ti contiennent des gènes impliqués à la fois dans les fonctions végétales et bactériennes.

Locus	Fonction	Plasmide Ti
<i>Vir</i> <i>Shi</i> <i>Roi</i>	Transfert d'ADN dans la plante Induction d'excroissances Induction de racines	Tous Tous Tous
<i>Nos</i> <i>Noc</i> <i>Ocs</i> <i>Occ</i>	Synthèse de la nopaline Catabolisme de la nopaline Synthèse de l'octopine Catabolisme	nopaline nopaline octopine octopine
<i>Tra</i> <i>Inc</i> <i>oriV</i>	Gènes de transfert bactériens Gènes d'incompatibilité Origine de réplication	Tous Tous Tous

parasite. Le principe est que *la cellule végétale transformée synthétise les opines que la bactérie peut utiliser*.

L'interaction entre *Agrobacterium* et une cellule végétale est illustrée dans la **Figure 33.14**. Le point important est le transfert par la bactérie d'une partie du plasmide Ti dans le noyau de la plante. La partie transférée du génome Ti est appelée **ADN-T**. Il est intégré dans le génome végétal, où il exprime les fonctions nécessaires à la synthèse des opines et à la transformation de la cellule végétale.

L'organisation des deux principaux types de plasmide Ti est illustrée dans la **Figure 33.15**. Environ 30 % des quelque 200 kb du génome Ti sont communes aux plasmides nopaline et octopine. Les régions communes incluent les gènes impliqués dans le transfert et l'intégration de l'ADN-T dans la plante, ceux requis pour l'induction des excroissances et des racines dans les plantes, et ceux nécessaires à la réplication et au transfert du plasmide dans les bactéries. Cependant, un réarrangement considérable des séquences s'est produit entre les plasmides.

La région T occupe environ 23 kb. Quelque 9 kb sont identiques dans les deux plasmides. Les régions T des plasmides nopaline et octopine portent des gènes différents pour le métabolisme des opines, mais partagent des séquences impliquées dans la régulation morphogénétique. Chaque plasmide porte des gènes nécessaires à la synthèse des opines (*Nos* ou *Ocs*) dans la région T ; les gènes correspondants pour le catabolisme des opines (*Noc* ou *Occ*) résident ailleurs sur le plasmide. Les fonctions morphogénétiques sont similaires, mais pas identiques, comme on peut le voir dans l'induction des différents types de tumeur.

La cartographie des fonctions génétiques du plasmide

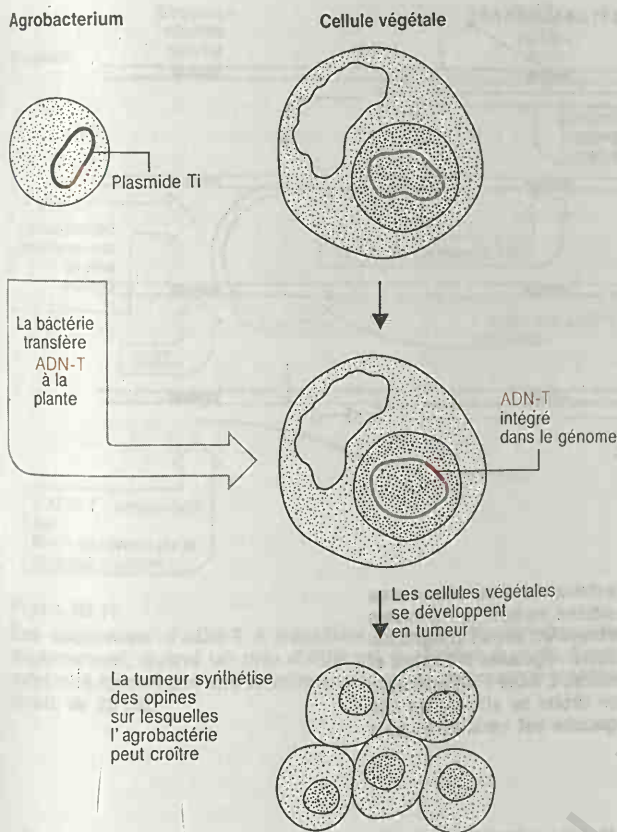


Figure 33.14
Un *Agrobacterium* contenant un plasmide Ti transfère l'ADN-T à une cellule végétale, où il s'intègre au génome et exprime des fonctions qui transforment l'hôte.

Ti n'est pas sans ambiguïté, étant donné les circonstances différentes dans lesquelles les différents gènes sont exprimés. Les cartes montrées dans la Figure 33.15 furent construites à l'aide de mutagenèse par insertions ou délétions, ce qui a permis de corréler les effets d'une mutation avec sa position sur la carte de restriction. Cette analyse a montré que les fonctions affectant l'oncogénicité — la capacité à former des tumeurs — ne sont pas limitées à la région T. Ces gènes, localisés en dehors de la région T, doivent être impliqués dans la mise en place de l'état tumoral, mais leurs produits ne sont pas nécessaires pour le perpétuer.

Les fonctions nécessaires pour initier l'oncogénicité sont communes aux plasmides nopaline et octopine. Elles peuvent être impliquées dans le transfert de l'ADN-T dans le noyau de la plante, ou dans des fonctions auxiliaires telles que l'équilibre des hormones végétales

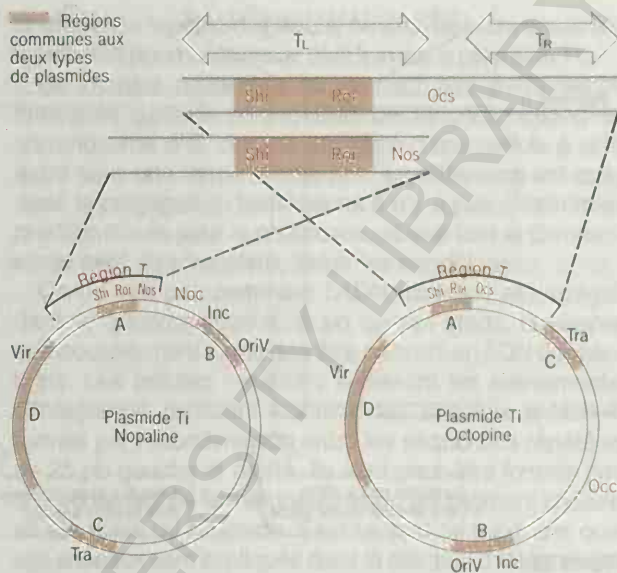


Figure 33.15
Les plasmides Ti nopaline et octopine contiennent une variété de gènes, incluant les régions T qui ont des fonctions chevauchantes.

dans le tissu infecté. Certaines des mutations sont spécifiques de l'hôte, empêchant la formation de tumeur chez certaines espèces végétales, mais pas chez d'autres.

Parmi de tels loci, les gènes de *virulence* identifient les fonctions impliquées dans le processus de transfert. Les gènes *virA-G* résident dans une région de 40 kb à l'extérieur de l'ADN-T et sont induits par des produits de faible poids moléculaire synthétisés par les cellules végétales qui sont susceptibles de se transformer parce qu'elles ont été blessées.

On peut diviser le processus de transformation en (au moins) deux étapes :

- Les *Agrobacteria* entrent en contact avec les cellules végétales et les gènes *vir* sont induits.
- Les produits des gènes *vir* permettent le transfert de l'ADN-T dans le noyau des cellules végétales, où il est intégré au génome.

L'ADN-T est-il intégré dans le génome végétal comme une unité intégrale ? Combien de copies sont intégrées ? Quels sites dans l'ADN végétal sont disponibles pour l'intégration ? Les gènes dans l'ADN-T sont-ils régulés exclusivement par des fonctions portées par le segment intégré ? Ces questions sont primordiales pour définir le processus par lequel le plasmide Ti transforme une cellule végétale en tumeur.

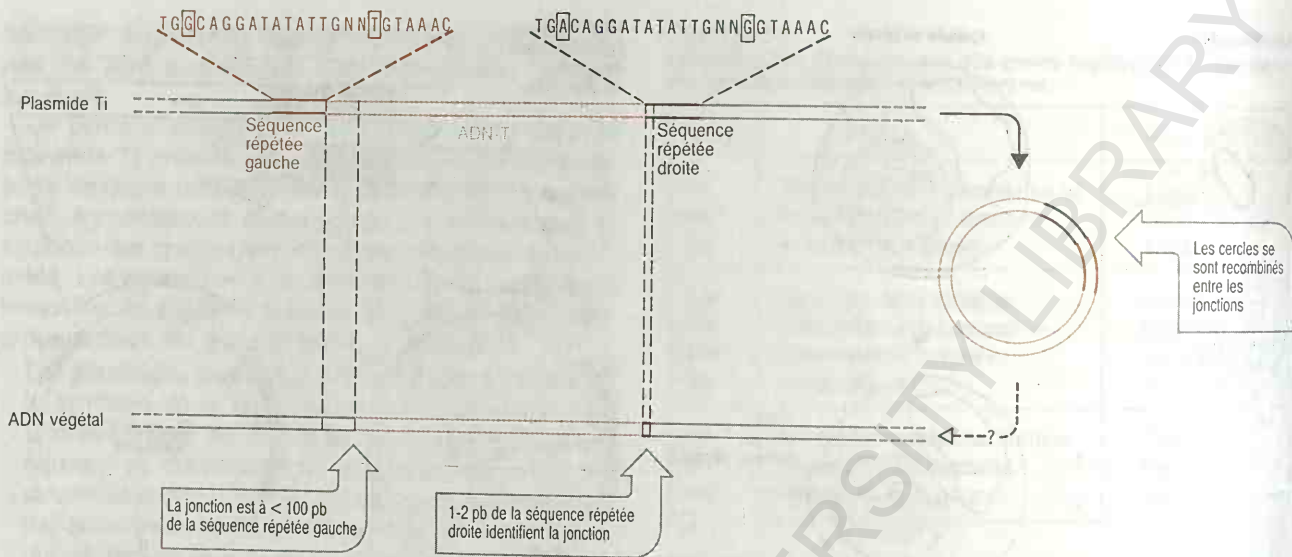


Figure 33.16

L'ADN-T a des séquences répétées continues de 25 pb près de ses extrémités. Les séquences répétées gauche et droite diffèrent en deux positions, et deux autres positions varient parmi toutes les séquences répétées. La séquence répétée droite est nécessaire pour le transfert et/ou l'intégration dans le génome végétal. Des cercles d'ADN-T sont formés dans les cellules végétales infectées et pourraient représenter un intermédiaire de la réaction de transfert/intégration. L'ADN-T intégré peut être trouvé sous la forme de copies simples ou en tandem. Sa jonction droite se situe à un site relativement précis au début des 25 pb répétées alors que la jonction gauche est variable et peut être distante de 100 pb de la répétition.

La bactérie n'entre pas dans la cellule végétale, mais transfère l'ADN plasmidique dans celle-ci. Est-ce que tout le plasmide Ti est transféré, après quoi l'ADN-T se trace un chemin vers le noyau ; ou est-ce que le processus de transfert lui-même est responsable de la sélection de la région T qui entre dans la plante ? La Figure 33.16 résume nos connaissances sur les structures de l'ADN impliquées dans les processus d'intégration et de transfert.

Une structure habituelle identifie les extrémités de la région T dans le plasmide Ti : des séquences répétées de 25 pb presque identiques démarquent l'ADN-T des régions bordantes. Cependant, contrairement à la situation observée pour les transposons, cette démarcation n'est pas précise : les limites de l'ADN-T intégré dans le génome végétal se trouvent proches, mais pas toujours à l'intérieur, de ces séquences répétées.

Les jonctions de l'ADN-T intégré avec l'ADN végétal ne sont pas spécifiques. Cependant, dans la majorité des insertions nopaline, l'ADN-T se termine au niveau de la séquence répétée de 25 pb située à l'extrémité droite. L'insertion est précise à un nucléotide près, et 1 ou 2 pb de la séquence répétée de droite sont incluses dans la séquence intégrée. Mais au niveau de la jonction gauche,

la variabilité est plus grande ; la limite de l'ADN-T dans le génome végétal peut être localisée au niveau de la séquence répétée de 25 pb ou à l'un quelconque d'une série de sites à 100 pb à l'intérieur de l'ADN-T.

L'ADN-T nopaline est intégré comme une unité intacte : ses régions internes sont maintenues en une séquence continue. Les plasmides octopine montrent plus de variation dans la sélection des régions de l'ADN-T intégrées. La Figure 33.15 délimite l'ADN-T octopine dans les régions T_L et T_R . La région T_L est toujours trouvée dans un génome végétal qui a reçu l'ADN-T. La région T_R est quelquefois, mais pas toujours, présente ; une réorganisation de la séquence trouvée dans cette région peut se produire. La relation existant entre les régions T_L et T_R dans le génome végétal n'est pas claire.

Plus d'une copie de l'ADN-T peuvent être intégrées à un site donné dans le génome végétal. Quand des copies en tandem sont présentes, des changements peuvent se produire dans les séquences proches des extrémités ; ils semblent impliquer une réorganisation des séquences du plasmide Ti au voisinage des extrémités. Ceci peut affecter les séquences situées de part et d'autre de l'élément répété de 25 pb dans le plasmide Ti d'origine. On ne sait pas si les copies en tandem sont formées *in*

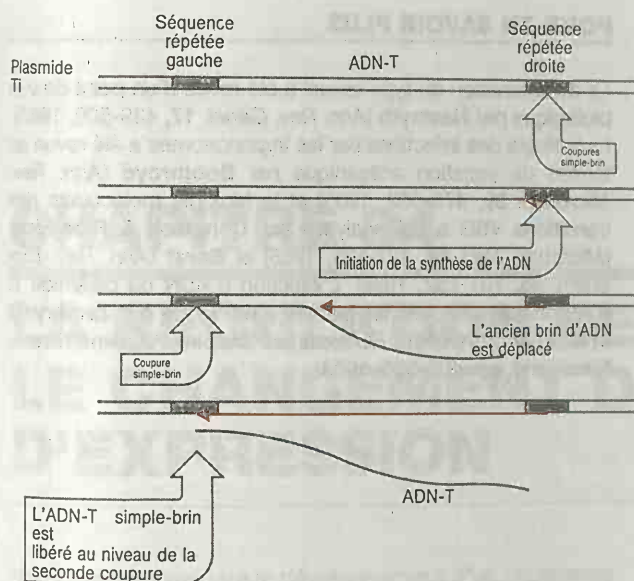


Figure 33.17

Les séquences d'ADN-T à transférer peuvent être formées par déplacement, quand un brin d'ADN est synthétisé à partir d'une extrémité formée par une entaille au niveau de la séquence répétée droite de 25 pb.

situ après qu'une seule copie ait été intégrée, ou si l'ADN-T lui-même engendre une extension en tandem qui est intégrée comme unité.

Seule la séquence répétée droite de 25 pb de l'ADN-T est nécessaire pour le processus d'intégration. De vastes délétions dans l'ADN-T ou à l'extrémité gauche ont peu d'effet, mais des délétions de la séquence répétée droite empêchent l'intégration. La séquence répétée fonctionne effectivement seulement dans son orientation d'origine, ce qui suggère que sa reconnaissance initie un processus de transfert polaire. On pourrait imaginer que l'ADN est clivé au niveau de ce site, et qu'ensuite la séquence située à gauche est reconnue pour les réactions de transfert ou d'intégration.

La Figure 33.17 illustre un modèle pour le processus de transfert. Une entaille est pratiquée au niveau de la séquence répétée de 25 pb droite. Elle fournit une extrémité amorce pour la synthèse d'un simple brin d'ADN. La synthèse du nouveau brin déplace l'ancien brin, qui est utilisé dans le processus de transfert. Le transfert est terminé quand la synthèse de l'ADN atteint une entaille située au niveau de la séquence répétée gauche. Ce modèle explique pourquoi la séquence répétée de droite est essentielle, et il explique la polarité du processus. Si

la séquence répétée de gauche ne peut pas être entaillée, le transfert pourra continuer plus loin sur le plasmide Ti.

Ce modèle ressemble étroitement aux événements impliqués dans la conjugaison bactérienne, quand le chromosome d'*E. coli* est transféré d'une cellule à une autre sous une forme simple brin. La différence est que, dans la conjugaison bactérienne il n'y a pas d'extrémité prédéterminée pour le processus, et que tout le chromosome peut être transféré, selon les conditions.

On ne sait pas comment l'ADN transféré est intégré dans le génome végétal. A un certain stade, le simple brin nouvellement formé doit être converti en ADN double-brins. Les cellules végétales subissant les événements d'intégration semblent contenir les anneaux d'ADN-T formés par recombinaison entre les séquences répétées de 25 pb gauche et droite. Ils sont peut-être formés par un événement de fermeture après qu'un fragment linéaire ait été libéré du plasmide. Il est tentant de supposer que ces anneaux sont impliqués dans la réaction d'intégration.

Quelle est la structure du site cible ? Les séquences flanquant l'ADN-T intégré ont tendance à être riches en paires de bases A-T (une caractéristique rencontrée chez certains éléments transposables). Les réarrangements de séquence qui se produisent aux extrémités de l'ADN intégré rendent l'analyse de la structure difficile. On ne sait pas si le processus d'intégration engendre de nouvelles séquences dans l'ADN cible, comparables aux séquences répétées cibles, produites au cours de la transposition.

L'ADN-T est exprimé à son site d'intégration. La région T_L contient plusieurs unités de transcription, chacune contenant probablement un gène exprimé à partir d'un promoteur individuel. Leurs fonctions concernent l'état de la cellule végétale, en maintenant ses propriétés tumorigènes, en contrôlant la formation des excroissances et des racines, et en supprimant la différenciation en d'autres tissus. Aucun de ces gènes n'est nécessaire pour le transfert de l'ADN-T.

Le plasmide Ti présente une organisation de fonctions intéressante. A l'extérieur de la région T, il porte des gènes nécessaires pour initier l'oncogenèse ; tout du moins certains ont un rapport avec le transfert de l'ADN-T, et on aimerait savoir si d'autres fonctionnent dans la cellule végétale pour influencer sur son comportement à ce stade. De même, à l'extérieur de la région T, il y a les gènes qui permettent à *Agrobacterium* de cataboliser l'opine que la cellule végétale transformée produira. A l'intérieur de la région T sont localisés les gènes qui contrôlent le stade transformé de la plante, ainsi que les gènes qui la conduisent à synthétiser les opines dont bénéficiera *Agrobacterium* qui a fourni l'ADN-T au départ.

D'une façon pratique, la capacité d'*Agrobacterium* à transférer l'ADN-T dans le génome végétal offre la perspective d'introduire de nouveaux gènes dans les plantes. Puisque les fonctions de transfert/intégration et d'oncogénèse sont séparées, il devrait être possible de construire de nouveaux plasmides Ti dans lesquels les fonctions oncogéniques ont été remplacées par d'autres gènes dont on espère tester l'effet sur la plante. L'existence d'un système naturel pour apporter des gènes au génome végétal devrait grandement faciliter et améliorer la manipulation génétique des plantes*.

POUR EN SAVOIR PLUS

La détermination du type sexuel a été revue d'un point de vue biologique par **Nasmyth** (*Ann. Rev. Genet.* **17**, 439-500, 1983). La biologie des infections par les trypanosomes a été revue en termes de variation antigénique par **Boothroyd** (*Ann. Rev. Microbiol.* **39**, 475-502, 1985) et la biologie moléculaire des transitions VSG a été analysée par **Donelson & Rice-Ficht** (*Microbiol. Rev.* **49**, 107-125, 1985) et **Borst** (*Ann. Rev. Biochem.* **55**, 701-732, 1986). L'injection d'ADN du plasmide Ti d'*Agrobacterium* dans les plantes a été revue par **Zambryski et al.**, in *Mobile Genetic Elements* (ed. Shapiro, Academic Press, New York), pp. 505-535, 1983.

* **N. d. T.** : Une telle méthode de création de dicotylédones transgéniques est aujourd'hui largement utilisée. Les plasmides Ti servant de vecteur transfèrent un ADN-T comportant les gènes à exprimer, les fonctions nécessaires à l'intégration, mais dépourvu des gènes nécessaires à la transformation.

CHAPITRE 34

LA RÉGULATION DES GÈNES : LE CHANGEMENT DES PROGRAMMES D'EXPRESSION

Nous avons considéré le développement d'un organisme adulte à partir d'un œuf fécondé, comme suivant une voie prédéterminée au cours de laquelle des gènes particuliers sont activés ou inhibés à des moments précis. Les conséquences d'une erreur dans ce processus peuvent être importantes : l'organisme peut se développer sous une forme anormale ou même mourir. De tels événements sont rares ; et puisqu'ils diminuent inévitablement la capacité d'un organisme à se perpétuer, ils sont contre-sélectionnés au cours de l'évolution.

L'existence des erreurs du développement fait partie du fardeau génétique, tribut de la capacité de l'appareil génétique à évoluer en offrant des variantes nouvelles et favorables à l'évolution. Et c'est l'existence dans le processus du développement, d'aberrations génétiques ou épigénétiques, qui donne un aperçu du réseau des interactions qui le contrôle : l'identification d'un mutant du développement permet de trouver les fonctions assurées par l'allèle sauvage.

D'un point de vue mécanistique, nous savons que l'élément clé de la régulation se situe au niveau du promoteur, et détermine si oui ou non un gène sera transcrit. Un gène procaryotique actif aura un promoteur accessible à l'ARN polymérase dont la liaison est parfois assistée par des protéines auxiliaires. Un promoteur eucaryotique actif aura une structure chromatiniennne modifiée, composée d'un site hypersensible sans nucléosome, pouvant nécessiter une activation en *cis* par un « enhancer ». Bien sûr, la transcription ne garantit pas

l'expression d'un gène, car des contrôles peuvent survenir lors des étapes ultérieures et modifier la stabilité ou la traductibilité chez les procaryotes, mais aussi la maturation et le transport chez les eucaryotes.

Des facteurs spécifiques pouvant être nécessaires à la reconnaissance des séquences du promoteur, ou à la régulation de l'activité d'un « enhancer », le produit d'un gène peut contrôler un autre gène. Quand des séries de tels événements sont connectées de telle sorte que l'expression du gène, à une étape donnée, contrôle l'expression d'autres gènes à l'étape ultérieure, une cascade de contrôles est formée. Les exemples les mieux connus de telles cascades sont ceux des bactériophages. Dans ce modèle, la caractéristique commune aux facteurs de régulation est que ce sont des protéines se liant à l'ADN (*DNA binding proteins*).

Nous supposons que le développement des eucaryotes utilise (au moins en partie) des gènes dont les produits fonctionnent d'une façon analogue : leur unique fonction est de réguler l'activité d'autres gènes. Une hypothèse souvent formulée est qu'ils le font directement, selon des modèles analogues à ceux discutés dans le Chapitre 27, dans lesquels une protéine régulatrice (ou un acide nucléique) se lie à un site présent dans la région du promoteur de chacun de ses gènes cibles. Nous avons quelques preuves de l'existence de tels sites ; nous avons moins d'informations concernant les facteurs qui peuvent agir sur eux *.

Notre principal objectif est d'identifier les gènes qui

* N. d. T. : Depuis 1985, plusieurs facteurs transcriptionnels eucaryotiques ont été en fait caractérisés et leur gène a parfois été cloné. Un phénomène très important a notamment été découvert : certains des facteurs de transcription spécifiques d'un type donné de différenciation ou d'une réponse particulière à un agent modulateur peuvent être en fait synthétisés de manière constitutive. Ils sont activés par un processus post-traductionnel dont le type serait la phosphorylation par différentes catégories de protéines kinases.

peuvent coder pour de tels facteurs, c'est-à-dire des gènes dont l'action est de réguler d'autres gènes. Un système de tels gènes est fourni par les loci homéotiques de la mouche du vinaigre, dans lesquels des mutations peuvent entraîner le développement d'une partie du corps à la place d'une autre. De ce fait, ils constituent d'excellents candidats pour être des gènes dont la fonction serait d'agir comme des « commutateurs » de régulation, analogues (par exemple) à celui qui intervient entre les cascades des cycles lytiques et lysogéniques de lambda. A l'origine décrits en termes de génétique classique, ces gènes ont maintenant été isolés et nous avons commencé de caractériser leur expression et leurs conséquences au niveau moléculaire.

LOCALISER LES MUTATIONS DANS LES EXONS ET DANS LES INTRONS

La caractérisation génétique des gènes régulateurs comme des autres gènes repose sur l'isolement de mutants. Au niveau génétique, cette approche est limitée par la nécessité de distinguer le phénotype mutant et (s'il est délétère) de propager les organismes mutants. Au niveau moléculaire, l'identification des mutants est affectée par la structure du gène. Un gène peut être une longue séquence d'ADN, comprenant des exons qui sont représentés dans l'ARN mature et des introns qui sont éliminés au cours de la maturation. Les mutations dans les exons qui affectent la séquence codante, peuvent être isolées grâce à leurs effets sur la protéine. Mais qu'en est-il du reste du gène ? Avant de considérer les tentatives pour caractériser les loci régulateurs, nous devons nous demander quel type de mutations nous sommes susceptibles de rencontrer et où elles seront localisées dans le gène.

L'étude systématique des mutations d'un gène chez les eucaryotes supérieurs n'a été réalisée que très rarement. Nous sommes trop limités par la nécessité d'identifier des mutants sur la base de modifications phénotypiques. L'une des études les plus détaillées concerne le locus *white* (blanc) de *D. melanogaster*, dans lequel une vaste série de mutations affectent à divers degrés, l'expression du gène.

Le locus *w* (*white*) fut identifié par la toute première mutation isolée chez la mouche du vinaigre, sous la forme d'un mâle dont les yeux apparaissaient délavés ou translucides et non de couleur rouge, comme le type sauvage. L'œil de type sauvage possède des pigments rouges et bruns sans relation structurale et synthétisés par des voies métaboliques différentes. Plusieurs gènes

ont été identifiés par leur capacité à interférer avec l'une ou l'autre de ces voies (les mutants *vermillion* (vermillon) n'ont pas de pigments bruns, les mutants *brown* (bruns) n'ont pas de pigments rouges, et ainsi de suite...).

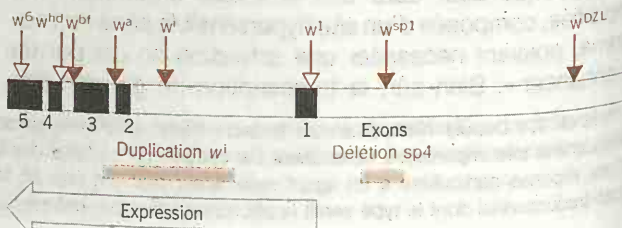
Les mutations du locus *w* font partie d'un même groupe de complémentation. Les produits des deux voies métaboliques de la synthèse des pigments ne sont pas présents dans les yeux des mutants *w*, ce qui suggère que le gène *w*⁺ ne code pas lui-même pour une enzyme impliquée dans la synthèse des pigments, mais régule directement ou indirectement la production ou le dépôt des pigments produits par les voies métaboliques rouge et brune.

La classification phénotypique des mutants *w* repose sur l'intensité de leur effet sur la formation des pigments. Les allèles nuls ou délavés — tel *w*¹, le premier allèle isolé — manquent complètement des pigments rouges et bruns. D'autres allèles ont des phénotypes intermédiaires dans lesquels des quantités réduites de l'un ou des deux pigments sont produites. Les allèles sont en général désignés par un exposant dont le nom reflète le phénotype. Par exemple, *w*^a est l'allèle abricot, *w*ⁱ l'allèle ivoire, *w*^e l'allèle éosine et ainsi de suite... La plupart des allèles affectent la qualité des pigments, mais le groupe *w*^{sp} affecte aussi leur distribution dans l'œil. Ceci est compatible avec l'idée que *w* gouverne le dépôt des pigments.

La Figure 34.1 représente une région d'environ 15 kb qui recouvre tous les sites connus des mutations *w*. Le gène comporte de 5' en 3', un exon connecté par un long intron à quatre exons qui sont séparés par de courts introns. Le transcrit mature est un ARNm de 2,6 kb, présent en très petite quantité (0,003 % des ARN polyadénylés) dans les cellules dans lesquelles *w* s'exprime. Le produit protéique est probablement un polypeptide hydrophobe de 78 000 daltons, dont la séquence suggère qu'il est peut-être localisé dans la membrane.

Figure 34.1

Le locus *white* peut comporter des délétions ou des insertions. Les exons sont en plein et les introns en vide. Les insertions sont indiquées par des flèches ; celles qui entraînent un phénotype nul sont indiquées par une flèche rouge en vide ; celles qui permettent la formation d'une quantité résiduelle de pigment sont indiquées par des flèches rouges en plein.



La plupart des mutations caractérisées à ce jour, résultent d'insertions de transposons (voir Chapitre 30). Leurs effets dépendent de leur localisation exonique ou intronique. Les mutations nulles sont créées par des insertions localisées dans les régions codantes qui empêchent la production d'une protéine active. Les mutations partielles sont localisées dans les introns, et sont supposées perturber à un certain degré l'expression, bien qu'une certaine quantité de protéine active soit produite. (La seule exception est w^l qui, tout en étant localisée dans un exon, est située dans la région 5' non codante.)

Nous comprenons la nature des mutations dans les exons, mais combien de fois une mutation localisée dans un intron pourra-t-elle être détectée par la perturbation de la fonction du gène ? Nous pouvons supposer que la plupart des modifications dans un intron passe inaperçue, puisque, après tout, ces séquences sont éliminées au cours de l'expression du gène. L'on peut supposer que les mutations dans les introns ne sont détectées que lorsqu'elles affectent la maturation des ARN en les empêchant, par exemple, d'être reconnus par le dispositif d'excision-épissage. Les mutations ponctuelles dont les effets moléculaires ont été caractérisés, changent habituellement les séquences consensus, empêchant l'excision correcte des introns (voir Chapitre 23). Des insertions dans le premier intron du locus w peuvent entraîner l'interruption du transcrit, modifier sa maturation ou sa stabilité. Leur relative bénignité est probablement due à la maturation correcte d'une minorité de transcrits.

L'idée que les insertions ont un effet indirect sur l'activité des gènes est renforcée par la nature des révertants. Un changement dans la fonction d'un locus est souvent associé à un changement de la séquence insérée. Par exemple, la délétion d'une petite partie de l'insertion w^l donne l'allèle w^e , qui a récupéré un certain degré d'expression des pigments. Les nombreux exemples de ce type d'effet impliquent que la perte d'une fonction n'est pas simplement due à l'insertion, mais peut être influencée, soit par la séquence, soit par la longueur du segment inséré.

La réversion ne se produit pas toujours par perte d'une quelconque partie du matériel inséré, comme cela est bien mis en relief par les dérivés de la mutation w^l . Contrairement aux autres modifications qui augmentent la taille du locus, w^l n'est pas due à l'insertion de matériel exogène, mais représente une duplication en tandem d'une région de 3 kb appartenant au locus w lui-même. Une réversion peut se produire directement au niveau de la mutation, ou par un changement compensateur secondaire.

Ainsi, le mutant peut retourner vers l'état normal en perdant exactement la duplication, probablement par recombinaison intrachromosomique, et ceci avec une fréquence relativement élevée d'environ 10^{-5} . Cependant, un révertant rare a été trouvé sous la forme du mutant w^c , qui a les yeux cramoisis ; il résulte de l'insertion d'une séquence de 11 kb à l'intérieur de la duplication w^l . L'allèle w^c est instable ; avec une fréquence d'environ 10^{-3} il peut retourner au type sauvage, au phénotype w^l ou à tout autre phénotype incluant des mutations stables ayant des délétions se terminant dans le locus w ; il semble possible que ces divers réarrangements soient contrôlés par la séquence insérée. Il paraît incroyable qu'une insertion aussi longue puisse restaurer le phénotype qui avait été supprimé par la duplication interne. Peut-être la séquence insérée place-t-elle un nouveau promoteur dans un site endommagé.

En amont du promoteur, l'on trouve deux sites au niveau desquels des mutations perturbent l'expression du gène w :

- Au niveau du site w^{sp} l'on trouve plusieurs insertions, délétions et des combinaisons des deux. En fait, les seules délétions détectables dans ce locus sont les allèles w^{sp3} et w^{sp4} , qui se produisent dans ou près du site de l'insertion w^{sp1} .
- L'insertion w^{DZL} est la plus éloignée vers la droite (environ 5 kb en amont).

Toutes les mutations qui touchent des fonctions régulatrices potentielles, telles les localisations spatiales et temporelles du dépôt des pigments, ou encore leur composition relative, sont localisées en amont du promoteur. Différents sites régulateurs correspondant à diverses fonctions régulatrices ont été identifiés par des expériences de transfection. Dans chaque tissu où le gène s'exprime, son activité est contrôlée de façon indépendante par des séquences localisées juste en amont du promoteur. Des séquences situées plus loin en amont sont nécessaires à son interaction avec un autre locus, *zeste*, qui contrôle une forme inhabituelle de régulation.

L'interaction *white-zeste* est un exemple de « **transvection** », un phénomène reposant sur l'appariement entre chromosomes. Il est mis en évidence par l'aptitude qu'ont les réarrangements entre chromosomes qui empêchent les synapses entre chromosomes homologues (comme on le voit dans les chromosome polytènes), de changer le phénotype alors même que le génotype dans son ensemble reste non altéré.

La base de la « transvection » repose sur la capacité d'un allèle à influencer l'activité de l'autre allèle *unique-ment lorsque les deux allèles sont maintenus accolés par synapse*. Ainsi, la « transvection » dépend probablement

de quelque effet structural exigeant le contact physique. Son existence laisse supposer que l'appariement des chromosomes (ou quelque effet qui lui est lié) se produit dans les cellules diploïdes normales.

La transvection peut produire un type sauvage encore plus sauvage, ou un type muté encore plus muté selon les systèmes. Au locus *BX-C*, la capacité qu'ont deux allèles mutants de produire un phénotype proche du phénotype normal peut dépendre d'une synapse. Au locus *w*, la transvection résulte d'une interaction avec le locus *zeste* (*z*), localisé à 0,75 unités de recombinaison à gauche de *w*, distance correspondant probablement à 300-400 kb. La mutation *z* est récessive. Elle n'a pas d'effet chez les mâles (qui ont seulement un chromosome X) ; chez les femelles homozygotes, elle réprime les allèles *w*⁺ associés en synapse, donnant une coloration pâle (*zeste*) des yeux. L'effet est supprimé si l'un des allèles *w*⁺ est transloqué ailleurs dans le génome, ou bien s'il est délété*.

En effet, la mutation *z* permet de détecter un état de diploïdie des allèles *w* : s'il y a synapse, leur expression est réprimée, mais s'ils ne sont pas en synapse, l'expression est normale. L'effet est indépendant de la distance entre *w* et *z*. Il peut être modifié par certains allèles de *w* obtenus par mutations dans la région 5' du locus.

La nature de l'interaction entre *z* et *w* suggère que, non seulement elle intervient sur des allèles en synapse, mais encore que l'appariement est nécessaire à l'apparition de l'effet. L'origine de la transvection est mystérieuse. Comme elle dépend du contact entre les allèles, l'hypothèse la plus simple est qu'elle est le résultat d'un effet de l'association en synapse sur l'expression des allèles.

Un autre exemple d'interaction dépendant d'une synapse est donné par la mutation *w*^{DZL} (*DZL* est une abréviation de « dominant *zeste*-like » : mutation dominante de type *zeste*). L'insertion *DZL* exerce une inhibition en *cis* de la fonction de la région *w* située sur sa gauche. Les caractéristiques très particulières de cet effet sont illustrées sur la Figure 34.2 ; l'allèle *w*^{DZL} empêche l'expression de l'allèle *w*⁺ qui entre en synapse avec lui. Il ne s'agit pas d'un effet classique de dominance en *trans* ; en effet, le même allèle *w*⁺ est exprimé normalement s'il est transloqué en un autre site où il ne peut plus entrer en synapse avec *w*^{DZL}.

L'allèle *w*^{DZL} a une insertion de 13 kb composée de

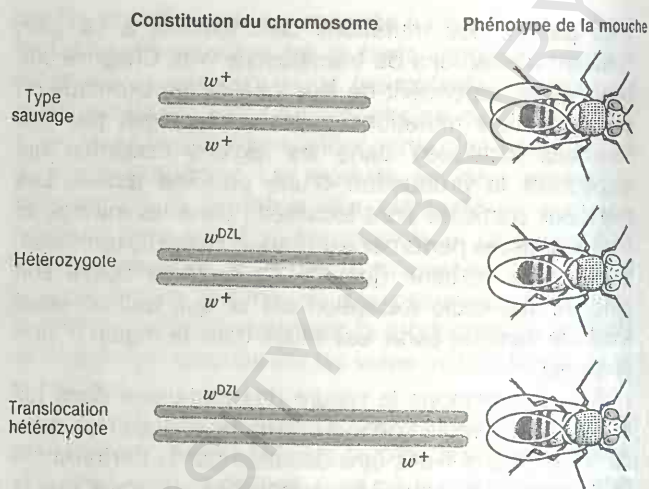


Figure 34.2

La mutation *w*^{DZL} empêche l'expression d'un allèle *w*⁺ en synapse avec lui mais ne peut pas agir sur un allèle *w*⁺ qui a été transloqué à un endroit différent.

deux séquences analogues mais non identiques (appelées éléments *FB*) entourant un segment d'ADN non homologue de 6 kb (séquence non répétitive du chromosome II). Comme une autre insertion du même type (en *w*^c), elle est instable, sa fréquence de réversion étant de 10⁻² à 10⁻³. Dans les révertants de type sauvage, une partie comportant toujours la région centrale de 6 kb est excisée. Cette excision partielle se produit probablement par recombinaison entre les séquences situées aux extrémités de l'insertion ; du fait de la présence de ces extrémités, les révertants restent instables. Cependant, puisque la présence d'une insertion de ce type à cet endroit n'entraîne pas l'apparition du phénotype *DZL*, c'est la séquence centrale de 6 kb plutôt que l'insertion en elle-même qui semble être responsable de l'effet de transvection.

L'effet de *DZL* est dû (probablement) à un promoteur porté par la partie centrale de l'insertion. La Figure 34.3 montre qu'un transcrit est initié dans la région *DZL* et s'étend d'un bout à l'autre du locus *w*. Il est épissé de telle sorte que la séquence *DZL* est connectée aux quatre exons de *w* situés en aval. La présence de ce transcrit s'accompagne d'une réduction du niveau de l'ARNm *w* de type sauvage. Cela suggère que l'activateur d'un

* N. d. T. : Il a été démontré très récemment (juin 1988) que le gène *z* codait pour une protéine se liant à l'ADN et pouvant agir sur la transcription. Quoique le mécanisme réel de la transvection demeure incertain, ce résultat suggère le schéma suivant : la protéine *zeste* se fixerait à un élément de contrôle situé à distance d'un des allèles *w*. Ce complexe nucléoprotéique distal interagirait avec les protéines fixées au niveau des régions de contrôle proximales du gène *w* de l'autre chromosome. En d'autres termes, des régions de contrôle distales d'un gène pourraient agir soit sur les éléments proximales du même allèle via la formation d'une boucle d'ADN, soit sur les éléments proximales de l'autre allèle grâce à la synapse entre les deux chromosomes.

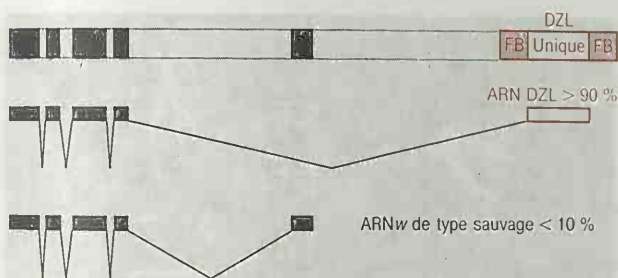


Figure 34.3

L'insertion *DZL* a un promoteur à partir duquel est initié un transcrit qui s'étend à travers le gène *white* ; l'expression à partir de ce promoteur réduit l'activité du promoteur *w*.

promoteur dans la région *DZL* inhibe le promoteur *w*. Le transcrit qui commence en *DZL* ne codant probablement pas pour une protéine *w*, il en résulte une réduction de la quantité du produit actif du gène.

Nous pouvons nous attendre à ce que le promoteur *DZL* ait une action en *cis* sur le promoteur *w*. Mais comment expliquer les effets de la transvection, quand la séquence *DZL* empêche l'expression en *trans* d'un allèle *w* associé par synapse avec un locus *white* situé en aval ? Un effet structural peut-il être transmis entre des chromosomes en synapse ? Le promoteur *DZL* peut-il entraîner que toute l'unité entrée en contact synaptique soit physiquement localisée en un endroit du noyau qui rend impossible l'activation du promoteur *white* ?

QUELQUES GÉNÉRALITÉS SUR LE DÉVELOPPEMENT DES *DROSOPHILES*

L'organisation générale des gènes d'eucaryotes est maintenant bien définie et il est admis qu'ils sont contrôlés par l'interaction entre des protéines et des éléments d'ADN situés le plus souvent en 5' des gènes ; nous attachons par conséquent, maintenant, une attention particulière aux relations entre l'expression des gènes et le développement des différentes parties d'un organisme. Nous aimerions savoir non seulement comment se fait la régulation chronologique de l'expression des gènes mais encore comment se fait la détermination spatiale ; comment un gène peut être activé dans une cellule, mais pas dans la cellule adjacente. Chez *D. melanogaster*, beaucoup des gènes impliqués dans l'établissement de la nature des structures du corps ont été identifiés. Avant de retourner au résultat du contrôle au niveau moléculaire, nous devons considérer les bases du développement de *Drosophila*.

La Figure 34.4 illustre le développement des structures du corps de *D. melanogaster*. L'embryogenèse dépend

de la mise en place des polarités selon les axes antéro-postérieur et dorso-ventral. Les mutations qui affectent la polarité peuvent entraîner les régions postérieures à se développer comme des régions antérieures, et les régions ventrales à prendre la forme des régions dorsales. Les gènes identifiés par ces mutations agissent très précocement ; certains ont été clonés, et, dans certains cas, une mutation peut être corrigée par l'injection de l'ARN de type sauvage.

Nous pouvons considérer le développement de *D. melanogaster* en termes de deux types d'unités : le **segment** et le **compartiment**. Leur relation n'est pas entièrement claire.

Le segment est une structure morphologique visible. La mouche adulte est formée d'une série de segments clairement délimités, et la larve a une série de segments correspondants séparés par des sillons. Nous nous intéresserons d'abord aux trois segments thoraciques (T) et aux huit segments abdominaux (A), dont le développement est presque entièrement connu.

Le compartiment est défini génétiquement comme une unité de développement. Dans l'embryon précoce, un nombre relativement petit de cellules fondatrices sont regroupées pour former un compartiment. Le destin de ces cellules est ensuite déterminé de telle sorte qu'aucune d'entre elles ni de leurs descendants ne peut bouger d'un compartiment à un autre. Dans l'embryon tardif et l'adulte, le compartiment peut être identifié comme comprenant uniquement des descendants des cellules fondatrices.

Un profil d'unités segmentaires est déterminé au stade blastoderme, quand une carte du destin (ou carte des territoires présomptifs) de l'embryon identifie les régions en termes de segments adultes qui se développeront à partir des descendants des cellules embryonnaires. Au stade blastoderme, la masse principale de l'embryon est divisée en une série de compartiments alternativement postérieurs (P) et antérieurs (A). Dans ces termes, un segment adulte est constitué d'un compartiment antérieur (A) suivi d'un compartiment postérieur (P).

Un autre type de division a été suggéré au niveau du blastoderme, lorsque des lignes de démarcation peuvent être vues entre des **parasegments**, chacun d'entre eux étant composé d'un compartiment P et d'un compartiment A qui est en fait situé sur son côté postérieur (voir Figure 34.4). Dans l'embryon de cinq à six heures, des sillons superficiels formés sur la surface de l'embryon séparent les parasegments adjacents.

Lorsque les **segments** se forment aux environs de la neuvième heure, les sillons s'approfondissent et peuvent se déplacer de telle sorte que chaque frontière entre les segments représente le centre d'un parasegment. Ainsi,

L'œuf

Les polarités sont établies le long de l'axe antéro-postérieur (tête-queue) et de l'axe dorso-ventral (dos-abdomen)

Blastoderme

Le destin de chaque région a été déterminé bien que les segments ne soient pas visibles

Embryons à 10 heures

Les segments thoraciques et abdominaux (et d'autres parties du corps non illustrées) se sont développés

Mouche adulte

Il y a 3 segments thoraciques et 8 segments abdominaux

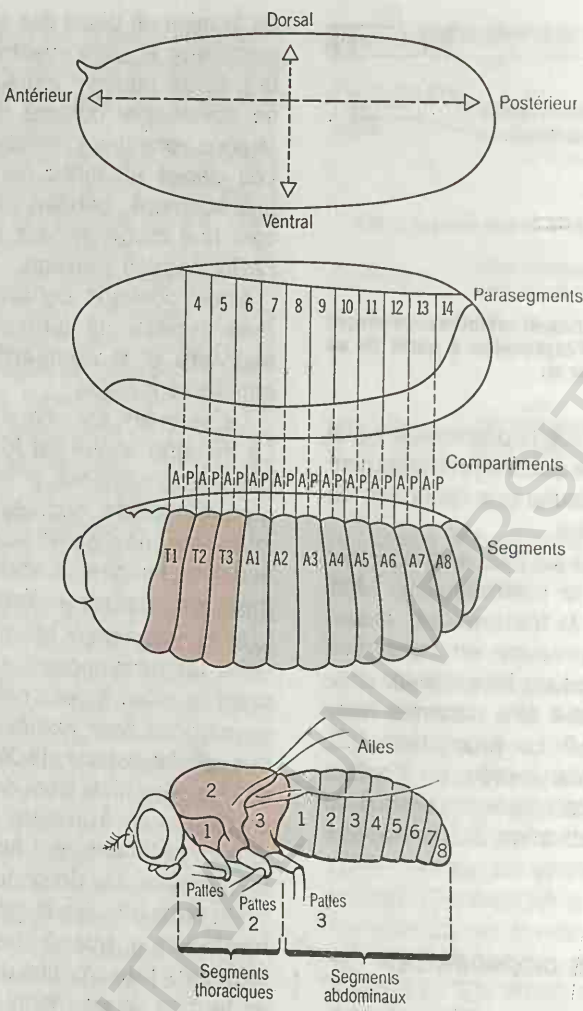


Figure 34.4
Le développement de *Drosophila* s'effectue grâce à la formation de compartiments qui forment des parasegments et des segments.

la partie antérieure du segment est dérivée d'un parasegment alors que la partie postérieure de ce segment est dérivée du parasegment suivant. Selon ce principe, le segment A3 a une partie antérieure (A) dérivée du parasegment 8 et une partie postérieure (P) dérivée du parasegment 9.

Des mutations qui affectent la structure des segments permettent d'identifier des loci particulièrement intéressants en ce qu'ils peuvent être impliqués dans la détermination de la nature des différentes parties du corps. Ces mutations permettent de définir les loci de segmentation et les loci homéotiques.

Les mutations de segmentation altèrent le nombre et la polarité des segments corporels. Elles peuvent être classées en fonction de la taille de l'unité de segmentation qu'elles affectent :

- Les mutants *pattern* (aspect général) ont perdu (habituellement) une partie du compartiment P de chaque segment, celle-ci ayant été remplacée par une duplication en miroir du compartiment A.
- Les mutants *pair rule* (règle de la paire) présentent des délétions équivalentes affectant un segment sur deux ; les segments modifiés peuvent avoir une numérotation paire ou impaire.
- Les mutants *gap* (brèche) ont des délétions intéressant un groupe de plusieurs segments adjacents.

Probablement moins de 20 loci sont impliqués dans la formation des segments. La nature générale des mutants de segmentation suggère que leur fonction est d'établir des règles qui président à la formation des segments ; une mutation peut changer une règle de telle sorte que

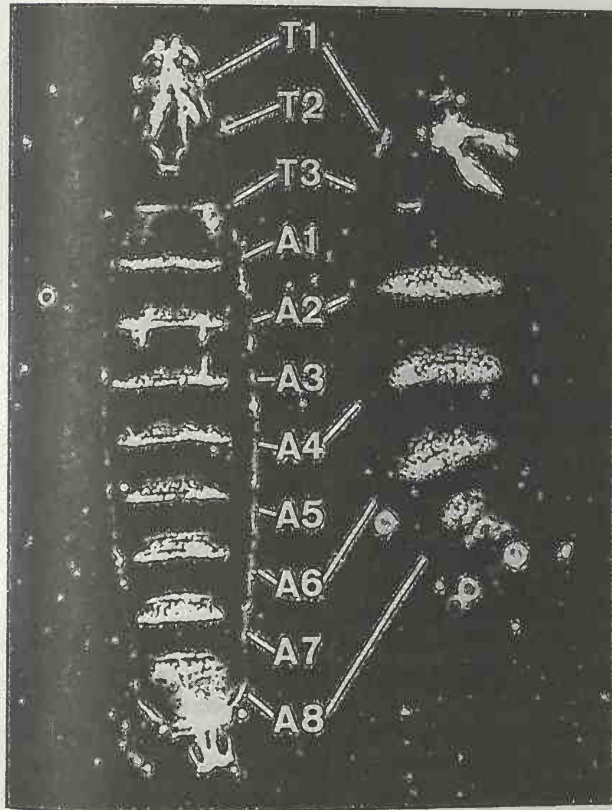


Figure 34.5
Les mutants *ftz* ont la moitié des segments présents chez le type sauvage.
Photographie aimablement fournie par Walter Gehring.

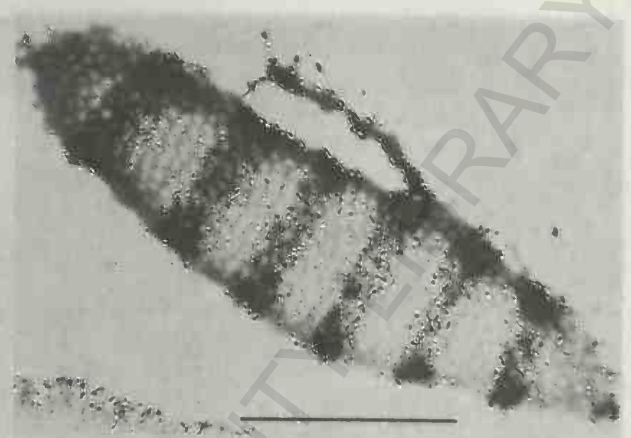


Figure 34.6
Les transcrits du gène *ftz** sont localisés en bandes correspondant aux parasegments pairs. Les régions exprimées chez le sauvage correspondent aux régions qui sont absentes chez le mutant *ftz* de la figure précédente.
Photographie aimablement fournie par Walter Gehring.

transcrits, localisés par hybridation *in situ* dans un blastoderme de type sauvage. Le gène est exprimé au niveau de sept bandes, chacune large de 3 à 4 cellules, tout au long de l'embryon. Les bandes correspondent aux parasegments ayant des nombres pairs (4 = T1P/T2A, 6 = T3P/A1A, 8 = A2P/A3A, etc.).

Cet aspect suggère ce que peut être la fonction du gène *ftz* : il doit être exprimé dans le blastoderme pour que se développent les structures qui dérivent des parasegments pairs. Les mutants ayant un gène *ftz* défectueux ne possèdent pas ces parasegments à cause de l'absence du produit du gène *ftz* durant la période au cours de laquelle ils doivent être formés.

Ceci peut correspondre à un modèle général de la fonction des gènes de segmentation. L'expression du gène dans une série de régions alternées contrôle le développement des compartiments. L'organisation des compartiments dans lesquels l'expression survient à un stade précoce détermine l'organisation des structures qui se développeront plus tard dans les larves et dans l'adulte.

L'organisation des compartiments dans lesquels les gènes de segmentation sont exprimés est extraordinairement précise. Une démonstration éclatante de cette précision est fournie par le gène *engrailed*. La fonction qui est codée par *engrailed* est nécessaire dans tous les segments et intervient dans la distinction entre

de nombreux segments (voire tous les segments) se forment imparfaitement. Les conséquences dramatiques des malformations des segments rendent ces mutants létaux au stade embryonnaire, ce qui signifie qu'ils meurent à des stades variés avant l'éclosion de l'organisme adulte.

Le diagramme général d'expression des gènes de segmentation renforce l'idée que *Drosophila* se développe par l'intermédiaire de compartiments. Le gène *fushi tarazu* (*ftz*) est un locus pair rule. La Figure 34.5 compare le diagramme de segmentation d'une larve de type sauvage et d'une larve *ftz* ; le mutant ne possède que la moitié des segments car un segment sur deux est absent.

Le gène *ftz* est constitué de deux exons, exprimés sous la forme d'un messenger de 1,9 kb qui est présent depuis le stade blastoderme jusqu'au stade gastrula du développement. La Figure 34.6 montre la localisation de ces

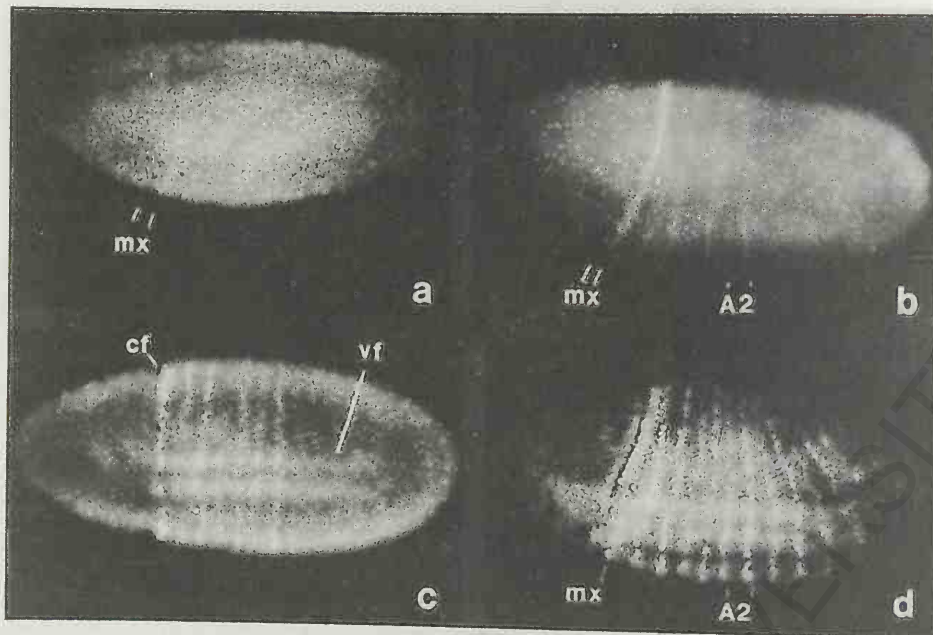


Figure 34.7

La protéine *engrailed* est localisée dans les noyaux et forme des bandes qui ont exactement une cellule en largeur.

Photographie aimablement fournie par Patrick O'Farrell.

les compartiments A et P. Les mutants de ce gène ne présentent, par conséquent, plus de distinction entre les régions antérieures et postérieures des segments.

Les anticorps contre la protéine codée par *engrailed* réagissent avec les noyaux des cellules qui l'expriment. Les régions dans lesquelles est exprimé *engrailed* forment l'aspect maintenant familièrement zébré en bandes adjacentes. Quand les bandes de protéine *engrailed* apparaissent pour la première fois, *elles n'ont une largeur que d'une cellule*. La **Figure 34.7** montre l'aspect observé à un tel stade, où chaque segment n'a une bande large que d'une seule cellule, les bandes commençant à s'élargir en plusieurs cellules.

En fait, l'organisation des bandes s'établit pendant une période d'environ 30 minutes, se déplaçant vers l'arrière de l'embryon. Initialement une seule bande est apparente ; ensuite un segment sur deux est ainsi marqué par une telle bande ; et finalement l'organisation complète comporte la présence d'une bande large de trois à quatre cellules correspondant à chaque segment.

Les **mutations homéotiques** entraînent le développement des cellules d'une lignée donnée en des cellules dont le phénotype appartient à une lignée différente. Certains mutants homéotiques transforment une partie d'un segment ou un segment entier en un autre type de

segment. Ils n'affectent pas le nombre, la taille ou la polarité des segments. Tout comme les gènes de segmentation, cependant, l'expression spatiale des gènes homéotiques correspond aux compartiments embryonnaires.

Les gènes homéotiques agissent durant l'embryogenèse. Nous nous intéresserons particulièrement au développement des structures larvaires appelées disques imaginaires, qui donnent naissance à beaucoup des structures adultes au moment de la métamorphose. Un mutant homéotique se comporte comme si les cellules d'une partie d'un disque imaginal avaient changé leur état de détermination, subissant de ce fait une métamorphose en la structure habituellement déterminée par les cellules d'un autre disque imaginal. Les transitions homéotiques peuvent ainsi entraîner le développement d'un segment de l'abdomen à la place d'un autre, le développement de pattes à la place d'antennes, ou d'ailes à la place d'yeux.

Des exemples particulièrement spectaculaires de mutations homéotiques apparaissent au niveau du locus *BX-C*, qui est impliqué dans le développement de l'abdomen. Un cas extrême convertit le troisième segment thoracique antérieur (qui porte les haltères, c'est-à-dire des ailes tronquées) en un tissu du type de celui du

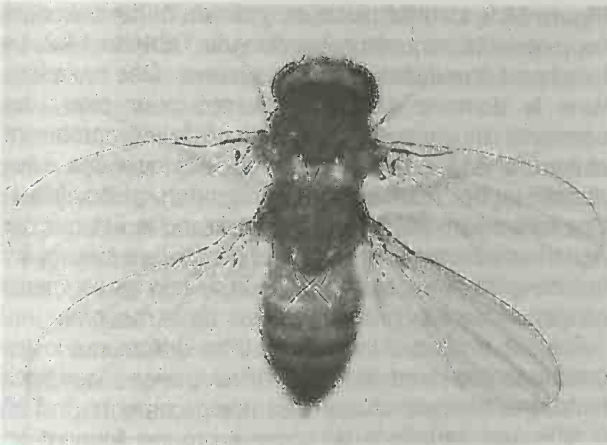


Figure 34.8
Une mutation dans *BX-C* produit une mouche avec 4 ailes.
(Cette mouche est un triple mutant dans *abx*, *bx*, et *pbx*).
Photographie aimablement fournie par Ed Lewis.

second segment (qui porte les ailes). Ceci crée une mouche qui est montrée dans la **Figure 34.8** avec 4 ailes au lieu de 2 comme normalement.

Des mutants homéotiques définissent des gènes qui contrôlent le programme de développement que devront suivre les descendants d'une cellule. Leur capacité à entraîner une commutation d'un programme en un autre suggère qu'ils identifient des gènes dont la fonction normale est de contrôler le développement. Il s'avère en fait que, non seulement les effets de ces gènes sur le développement sont très intéressants, mais aussi que l'organisation moléculaire de ces loci (au moins certains d'entre eux) est inhabituelle et pose elle-même d'étranges questions à propos de l'expression des gènes.

LES LOCI COMPLEXES SONT EXTRÊMEMENT GRANDS ET INTERVIENNENT DANS LA RÉGULATION

La plupart des gènes de *D. melanogaster* sont relativement petits ; l'exemple de *white* donne quelques indications sur la nature des mutations et la complexité des éléments de régulation. Mais certains loci se comportent de manière différente, qu'il est difficile de concilier avec la notion que chacun contient un gène unique codant pour un produit particulier. Ils sont appelés **loci complexes**.

Un gène conventionnel — même un gène mosaïque — est identifié au niveau de la carte génétique par un groupe étroitement lié de mutations qui ne complètent pas. Une caractéristique des loci complexes est la présence de groupes de mutations relativement bien espacés, s'étendant sur une grande distance de carte génétique, et montrant souvent un diagramme complexe de complémentation. Les mutations, par exemple, peuvent se regrouper en plusieurs groupes chevauchants de complémentation. Les mutations individuelles peuvent avoir des effets morphologiques différents et complexes sur le phénotype. Au niveau moléculaire, certains loci complexes sont très grands. Le **Tableau 34.1** résume l'organisation de certains de ceux qui ont été caractérisés.

Une particularité des loci complexes est leur très grande taille et le fait que les mutations peuvent survenir n'importe où dans une grande région génomique ; leur organisation génétique peut cependant prendre plusieurs formes :

- *BX-C* est le cas classique dans lequel une unité d'expression commune contient plusieurs gènes.
- *achaete-scute* possède de multiples gènes qui sont exprimés séparément, apparemment sous la forme

Tableau 34.1
Les loci complexes chez *D. melanogaster* comprennent les plus grandes unités fonctionnelles connues.

Locus	Partie du corps affectée	Longueur du Locus	Taille des ARNm connus	Nombre de Gènes	Structure des Gènes
<i>BX-C</i>	thorax/abdomen	> 300 kb	3,2 & 4,3 kb	> 3	interrompue
<i>Antp</i>	tête/thorax	> 100 kb	3,2-4,8 kb	1	interrompue
<i>Notch</i>	neurones	> 37 kb	11,7 kb	1	9 exons
<i>Achaete-scute</i>	soies	> 75 kb	1,2 & 1,6 kb	3	non connue

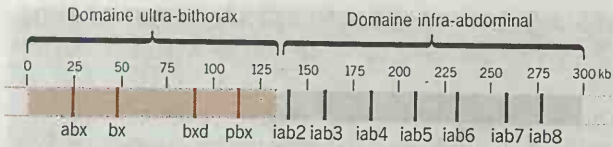


Figure 34.9
Le locus *BX-C* possède deux domaines, chacun d'entre eux comprenant plusieurs types de mutations homéotiques.

d'unités de transcription relativement petites ; cependant, les phénotypes mutants produits par des lésions de l'ADN de toute l'unité (notamment d'importantes régions non transcrites) sont identiques.

- *Notch* comporte un seul gène contrôlé au niveau de la transcription et des réactions d'excision épissage ; il montre cependant des profils de complémentation inhabituels.

Certains des loci complexes interviennent dans le contrôle du développement du corps de l'insecte adulte. Le locus homéotique complexe le plus classique est *BX-C*, le **complexe bithorax**, qui est caractérisé par plusieurs groupes de mutations homéotiques qui affectent le développement du thorax et causent des modifications morphologiques majeures au niveau de l'abdomen. Quand la totalité du complexe est délétée, les insectes meurent tardivement au cours du développement embryonnaire. Dans le complexe, cependant, peuvent également survenir des mutations qui sont viables mais qui changent le phénotype de certains segments.

La carte génétique de *BX-C*, qui est donnée dans la

Figure 34.9, identifie plusieurs types de mutations, dont les propriétés sont résumées dans le **Tableau 34.2**. Le locus peut être divisé en deux domaines. Des mutations dans le domaine *Ultra-bithorax* ont pour cibles les segments thoraciques T2 et T3 et le compartiment abdominal A1A. Des mutations dans le domaine *infra-abdominal* ont pour cibles les segments abdominaux. Une caractéristique fondamentale du locus est que les mutations affectant des segments particuliers siègent sur la carte génétique dans le même ordre que les segments correspondants au niveau du corps de la mouche.

C'est à propos du domaine *Ultra-bithorax* que les connaissances sont les plus développées ; certaines mutations affectent chacun des compartiments de T2P à A1A. Les mutations de chaque groupe forment un ensemble d'allèles mutants ayant un phénotype similaire, dont les intensités varient mais qui présentent tous le même type de transformation tissulaire. Chaque ensemble identifie un gène qui agit sur un compartiment particulier ou sur une paire de compartiments.

Toutes les mutations individuelles de *BX-C* convertissent le compartiment cible de telle sorte qu'il se développe comme s'il était localisé à la position correspondante du segment précédent. Par exemple, *pbx* cause le développement de T3P comme s'il s'agissait de T2P. Plusieurs fonctions de ce type sont connues dans le domaine *Ultra-bithorax*, cela s'appliquant également aux domaines *infra-abdominaux*. (Quoique nous ayons plus d'informations à propos de la région qui agit sur les trois loci thoraciques, il devrait y avoir un ensemble de fonctions plus important intervenant au niveau des régions

Tableau 34.2
Chaque mutation récessive dans *BX-C* entraîne un phénotype caractéristique.

Mutation	Nom	Segment affecté	Phénotype
<i>abx</i>	<i>antéro-bithorax</i>	T2P/T3A	converti en T2
<i>bx</i>	<i>bithorax</i>	T3A	converti en T2A
<i>bxd</i>	<i>bithoraxoïde</i>	A1A	converti en T3A
<i>pbx</i>	<i>post-bithorax</i>	T3P	converti en T2P
<i>Ubx</i>	<i>Ultra-bithorax</i>	T3	somme de <i>abx</i> , <i>bx</i> , <i>bxd</i> , <i>pbx</i>
<i>iab2</i>	<i>infra-abdominal</i>	A2A	converti en A1
<i>iab3</i>	»	A3-A6	converti en A2
<i>iab4</i>	»	A4	converti en A3
<i>iab5</i>	»	A5	converti en A4
<i>iab6</i>	»	A6	converti en A5
<i>iab7</i>	»	A5-A8	converti en A3-A4
<i>iab8</i>	»	A8	converti en A7

abdominales.) Dans le complexe *BX-C* pris dans son ensemble il y a, ainsi, des fonctions individuelles qui jouent des rôles analogues dans le développement des compartiments successifs de la mouche.

La méthode de la marche sur le chromosome (chromosome walking) a permis de construire une carte de plus de 300 kb couvrant toute la région *BX-C*. La **Figure 34.10** indique la localisation des différents types de mutations.

Des mutations récessives, représentant une perte de fonction, ont une organisation surprenante. Des mutations de type *bx* et *bxd* couvrent chacune une région importante mais bien délimitée. Les mutations *abx* et *pbx* sont causées par des délétions. L'ordre de ces mutations sur le génome est le même que l'ordre des segments qu'elles affectent au niveau de l'abdomen de la mouche.

Le type *Ubx* se comporte comme s'il possédait plusieurs des mutations de fonctions individuelles identifiées par ailleurs grâce aux mutations récessives du domaine *Ultra-bithorax*. La plupart des allèles *Ubx* ont des réarrangements cytologiquement visibles du chromosome qui, au niveau moléculaire, se révèlent être des délétions éliminant des longueurs variées de la partie gauche du locus. Une notion importante est qu'une brèche, dans le domaine peut suffire à éliminer ces fonctions. Certaines autres mutations *Ubx* correspondent en effet à de petites délétions en différents sites du locus.

L'expression de très grands loci est difficile à analyser. Il est extrêmement hasardeux d'identifier des molécules

d'ARN d'environ 100 kb de long ; et de toute façon, si l'épissage apparaît avant que la transcription ne soit achevée, les transcrits primaires intacts peuvent très bien ne pas exister. Des molécules d'ARN ayant des profils d'épissage différents pourraient être présentes à faible concentration, rendant bien difficile l'identification de chacune de ces formes. Les données actuelles sont ainsi quelque peu préliminaires.

Une carte de transcription de la région *Ultra-bithorax* est présentée sur la **Figure 34.11**. Il y a deux unités de transcription. L'unité *bxd* d'environ 25 kb s'étend depuis les mutations *pbx* jusqu'à y compris les mutations *bxd*. L'unité *Ubx* d'environ 75 kb couvre les mutations *bx* et *abx*. L'une et l'autre de ces unités possèdent des propriétés extrêmement curieuses.

La grande unité de transcription *Ubx* donne naissance à plusieurs petits ARN. Une espèce transitoire de 4,7 kb apparaît la première, puis est remplacée par des ARN de 3,2 et 4,3 kb. Une caractéristique commune à ces deux ARN tardifs est qu'ils comprennent l'un et l'autre des séquences dérivées des deux extrémités du transcrit primaire. (Naturellement, il se pourrait qu'existent d'autres ARN non encore identifiés.)

Nous n'avons pas encore d'idées précises quant aux fonctions de la protéine pour laquelle codent ces ARN ; nous savons que les premiers et les derniers exons sont très grands alors que les exons internes sont plutôt petits. De petits exons appartenant à cette grande unité de

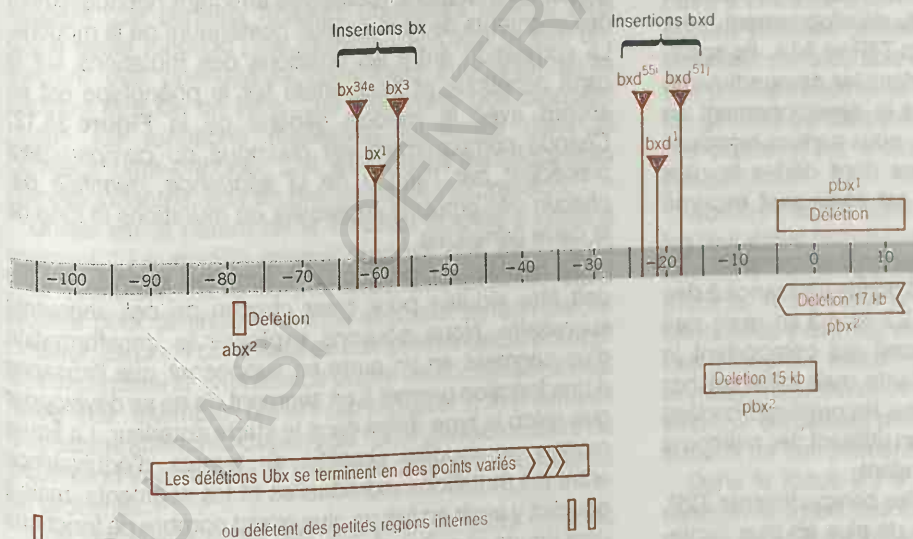


Figure 34.10
La plupart des mutations dans le complexe *bithorax* sont la conséquence d'insertions ou de délétions qui sont localisées en groupes selon la fonction touchée.

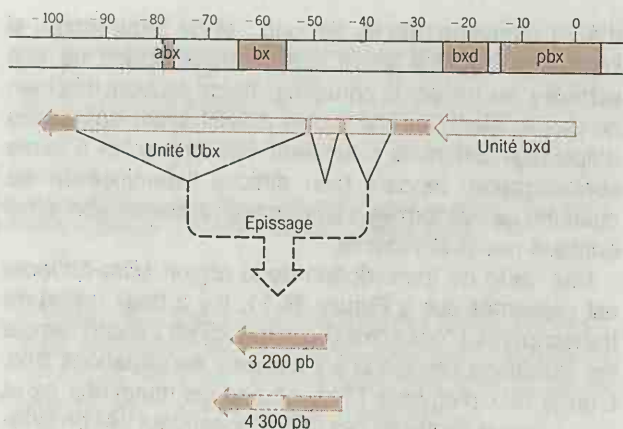


Figure 34.11

Le domaine *Ultra-bithorax* est transcrit en deux grandes unités ; *bxd* est non codant et *Ubx* est épissé de façon alternative pour donner des ARNm qui codent pour des protéines ayant des régions N-terminales apparentées.

transcription peuvent être intégrés dans le messager final grâce à des phénomènes d'épissage alternatif. Une telle organisation pourrait en partie expliquer l'existence de groupes de complémentation chevauchants, dans lesquels des mutations apparaissent parfois appartenir à des gènes indépendants dans le locus et parfois faire partie d'un locus complexe unique.

Lorsque l'expression de la région *Ultra-bithorax* (*Ubx*) est initialement détectée aux environs du stade blastoderme, des transcrits sont trouvés dans une région qui correspond au parasegment 6. Ils demeurent concentrés dans cette région tout au long du développement alors qu'elle devient les compartiments T3P et A1A. Ils apparaissent également dans T2P et dans les compartiments antérieurs de A2 à A7 alors que le développement se poursuit. (Dans cette description, nous parlons principalement des tissus ectodermiques dont dérive la plus grande partie de l'insecte ; *Ubx* est également exprimé dans la couche mésodermique.)

Un modèle de travail de l'expression des transcrits *Ubx* suppose qu'un épissage alternatif donne naissance à des ARN (incluant des espèces de 3,2 et 4,3 kb mais pas uniquement limités à ces espèces) qui comportent le début et la fin de la région codante mais diffèrent par leurs exons internes. Sur cette base, les protéines codées par cette unité ont été localisées en utilisant des anticorps dirigés contre le premier exon codant.

Les anticorps réagissent avec les compartiments T2P, T3A (faiblement), T3P, A1A, puis de plus en plus faiblement avec les segments de A2 à A8. Aucune coloration n'est trouvée dans les larves qui ont, à l'état homozygote, une petite délétion *Ubx* localisée proche de l'extrémité

5' du transcrit, et introduisant un codon de terminaison prématuré dans la séquence codante. De ces résultats, nous pouvons conclure que l'unité *Ubx* code pour probablement toute une famille de protéines qui sont concentrées dans les compartiments affectés par les mutations *Ultra-bithorax*. Les protéines sont localisées dans le noyau, ce qui est cohérent avec l'idée qu'elles sont des facteurs de régulation de l'expression des gènes.

Des mutations dans *bxd* modifient la distribution de la protéine, réduisant la réaction avec les compartiments T3P et A1A. Par conséquent *bxd* contrôle *Ubx*. En dépit de sa grande taille, le transcrit *bxd* ne semble pas coder pour des protéines. Nous savons par analyse génétique que *bxd* agit en *cis* sur *Ubx*. Nous ne savons pas si les effets nécessitent la transcription de *bxd* ; nous n'avons, en réalité, aucune idée de la manière dont le transcrit *bxd* de 25 kb pourrait contrôler en *cis* le transcrit adjacent *Ubx* de 75 kb.

La moitié droite de la région *BX-C* est le domaine *infra-abdominal* qui contrôle le développement des segments de A3 à A8. Tout comme le domaine *Ultra-bithorax*, il contient plusieurs groupes de mutations qui altèrent des segments successifs de la mouche. Il existe un profil complexe de complémentation entre ces mutations, suggérant qu'interviennent des interactions agissant en *cis* et d'autres agissant en *trans*. La fonction est perdue si *iab2* est séparé par coupure de *iab3,4,5* ou si *iab7* est séparé de *iab5,6*.

En se déplaçant le long du gène de la gauche vers la droite, les mutations récessives affectent habituellement des segments de plus en plus postérieurs de la mouche. La corrélation entre les positions des mutations sur la carte génétique et leurs effets sur le phénotype est en accord avec le modèle général de la Figure 34.12. Chaque compartiment est distingué du compartiment précédent par l'action de la séquence identifiée par chacun des groupes successifs de mutations le long de la carte génétique.

Ce modèle suppose qu'une fonction supplémentaire doit être ajoutée pour créer chacun de ces segments successifs. Nous pouvons expliquer la transformation d'un segment en un autre en supposant que l'absence d'une fonction permet à un segment de ne se développer que selon le type défini dans le stade antérieur. La figure montre un modèle simplifié en termes d'équivalence entre les mutations individuelles et les segments, mais il pourrait y avoir en fait un plus grand nombre de fonctions génétiques et aussi d'unités comportementales.

En principe *abx* peut être nécessaire pour le développement de T2 ; si cette fonction est absente le segment se développe à la place comme T1. De la même manière,

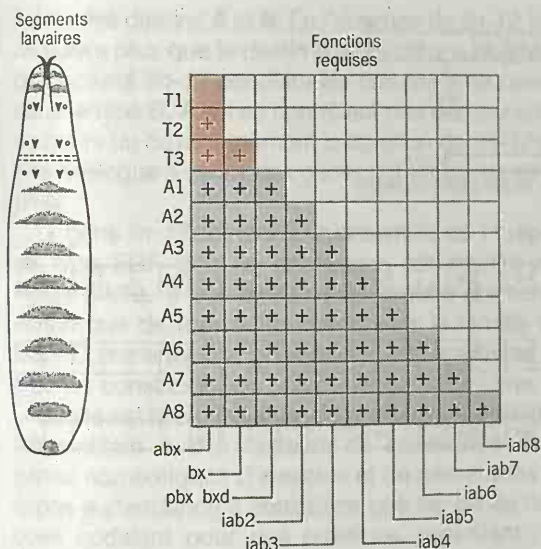


Figure 34.12
La région BX-C pourrait définir des segments thoraciques et abdominaux si la fonction des loci successifs dans la région était nécessaire pour distinguer chaque segment du précédent.

iab2 est, plus tard, nécessaire au développement de A2 ; l'absence de cette fonction conduit le segment à se développer comme A1. Ceci est en un certain sens un modèle binaire dans lequel une commutation supplémentaire doit être actionnée à chaque segment.

La structure complexe de l'unité de transcription *Ubx* suggère que certains mutants pourraient affecter plus d'une seule fonction ; les mutants dans *abx* et *bxd* ont souvent des phénotypes complexes suggérant qu'ils peuvent être de ce type. Les mutations *Ubx* bloquent toutes les fonctions récessives localisées dans le domaine *Ultra-bithorax*, soit en déléant les séquences correspondantes, soit en prévenant leur expression.

Quelle est la nature de la fonction ajoutée à chaque segment ? Le modèle original propose que *BX-C* code pour toute une série de protéines, une protéine étant ajoutée pour distinguer chaque segment de son prédécesseur. Le profil complexe de complémentation pourrait être expliqué si la synthèse de protéines successives était reliée par exemple par l'intermédiaire d'événements d'épissage alternatif. Cependant l'analyse moléculaire du locus rend improbable la synthèse d'un groupe distinct de protéines.

Une possibilité serait que des changements minimes dans le produit protéique accomplissent un ajustement fin dans lequel son action sur le segment serait modifiée. Cependant au moins certains effets doivent utiliser d'autres voies comme le suggère l'absence de toute fonction

codante apparente dans les régions identifiées par les mutations *bxd* et *pbx*.

Le locus *BX-C* est le plus grand et le plus complexe des loci de ce type, mais *ANT-C* est lui aussi très grand. On considère souvent qu'il contient plusieurs gènes homéotiques, notamment *proboscipedia* (*pb*), *Sex combs reduced* (*Sxr*) et *Antennapedia* (*Antp*). Le gène *Antp* donne son nom au complexe, et parmi les mutations qui l'intéressent il y a des allèles qui changent les antennes en seconde patte ou bien les seconde et troisième pattes en premières pattes. A nouveau le locus est très grand au niveau moléculaire. *Antp* couvre plus de 100 kb mais apparemment ne donne naissance qu'à une seule protéine de 48 000 daltons. Les rapports de *pb* et de *Sxr* à *Antp* ne sont pas clairs.

Le locus a les caractéristiques d'un gène mosaïque puisque des régions fort distantes se retrouvent dans les sections contiguës de molécules d'ARN beaucoup plus petites, de moins de 5 kb de long. L'analyse complète des transcrits reste à faire.

La différence entre les loci complexes et les gènes ordinaires de *D. melanogaster* est tout à fait spectaculaire. La plupart des gènes de la mouche du vinaigre, qui ont été caractérisés, sont relativement petits. Soit ils n'ont pas d'introns, soit leurs introns sont de taille réduite. En revanche les loci complexes sont immenses et sont épiés en des espèces de molécules d'ARN d'un ordre de grandeur plus petit. Pourquoi les loci qui interviennent dans le contrôle du développement de la larve embryonnaire à l'insecte adulte sont-ils différents des gènes qui codent pour les protéines habituelles de l'organisme ? Leur taille considérable est-elle nécessaire à la production des produits alternatifs ? Y a-t-il une relation entre la taille du gène et certains mécanismes temporels déterminés, par exemple, par le temps que cela prend de transcrire l'unité ?

Notch (*N*) est un autre locus complexe au niveau duquel les mutations ont plusieurs effets sur le phénotype apparemment sans rapport les uns avec les autres. Les mutants *N* donnent un phénotype dominant aux ailes encochées (notched) chez les hétérozygotes *N/+*, et sont des récessifs létaux comme le montre la mort des homozygotes *N/N* chez lesquels s'accumulent un nombre excessif de cellules précurseurs de neurones. Les mutations *N* ne complètent pas l'une l'autre.

Dans le locus il y a plusieurs groupes de mutations récessives ayant des effets différents sur le phénotype. Certaines altèrent la formation des ailes ; d'autres affectent les yeux ou les soies. Ces mutations sont toutes définies comme faisant partie du groupe *Notch* du fait de leur incapacité à compléter les mutations *N*, mais,

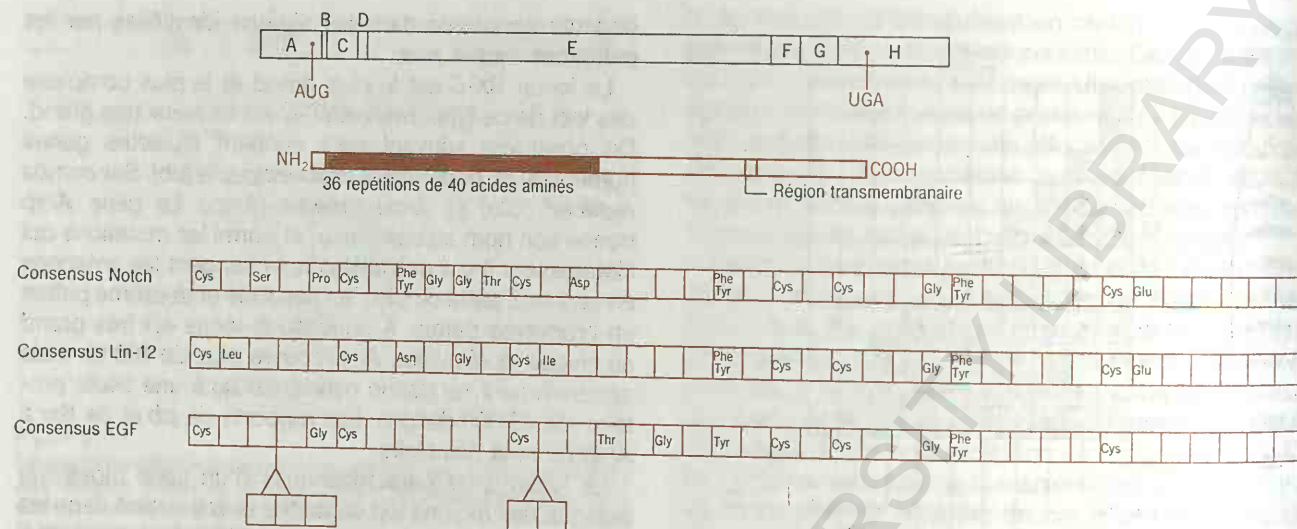


Figure 34.13

La protéine *Notch* possède plusieurs domaines, l'un d'entre eux comprenant une série de répétitions proches de EGF.

lorsqu'elles sont testées parmi elles-mêmes, elles peuvent être groupées en au moins quatre groupes de complémentation.

Le locus *Notch* contient une unité de transcription qui s'étend sur 37 kb et comporte 9 exons dont la taille va de 130 pb à 7 250 pb. L'ARN correspondant est long de 10,5 kb. La plupart des mutations *N* sont regroupées dans les exons ; certaines des mutations récessives représentent des insertions dans les introns. Quand nous pourrions caractériser les mutations en déterminant leurs effets sur la séquence d'ADN du locus, nous serons alors capables de comprendre la cause de cette nature complexe des mutations *Notch*.

La séquence de l'ARNm *Notch* de type sauvage code pour une protéine de 2 703 acides aminés. La Figure 34.13 illustre son organisation générale. La caractéristique la plus remarquable réside en un segment qui est codé par les segments B, C, D et une partie de E qui comprend 36 répétitions d'une séquence d'environ 40 amino-acides. L'unité de répétition est riche en cystéine et en fonction du critère de l'intervalle entre les résidus cystéines, elle semble reliée à l'EGF de mammière (Epidermal Growth Factor, facteur de croissance des cellules épidermiques).

La fonction d'EGF n'est pas complètement élucidée mais l'on sait qu'il s'agit de l'un des nombreux facteurs de croissance qui semblent agir en stimulant la prolifération de certaines cellules. C'est un polypeptide de 53 amino-acides chez la souris et chez l'homme ; dans

l'un et l'autre cas, il provient du clivage d'un précurseur polypeptidique d'environ 1 200 amino-acides qui contient 10 copies de l'unité de répétition. Plusieurs protéines comportent des séquences ayant la série caractéristique de 6 cystéines et pour le moment nous ne connaissons pas le lien qui pourrait être commun à leur fonction. Ces protéines sont, sur la base de leur cystéine décrite, comme appartenant à la famille ressemblant à EGF (EGF like).

Nous n'en savons pas assez sur la fonction des répétitions de type EGF pour prédire leur fonction. Nous savons que les 6 résidus cystéines forment trois paires de ponts di-sulfure qui peuvent donner à cette région une structure caractéristique. Ces molécules lient-elles un facteur commun ? Ce facteur permet-il d'identifier une fonction protéique commune ou bien est-il plutôt un cofacteur requis dans de nombreux types différents de fonctions protéiques ?

Les relations entre *Notch* et EGF pourraient être considérées comme une curiosité mais en fait elles sont dotées d'une signification renforcée par l'observation de relations similaires entre le gène de nématode *lin-12* et EGF. Le locus *lin-12* intervient dans le contrôle du développement des lignées cellulaires de *C. elegans*, (*lin*, ainsi nommée pour lignage).

Des mutations récessives qui causent une perte d'activité de *lin-12* et des mutations dominantes élevant en apparence cette activité ont des effets opposés sur le destin cellulaire. Une paire de cellules peut habituellement

suivre des destins **A** et **B**. En l'absence de *lin-12*, la cellule ne suivra plus que le destin **A** alors qu'une augmentation de l'activité *lin-12* conduira les cellules à se développer dans le type **B**. Ainsi en contrôlant des décisions binaires au cours du développement la fonction de *lin-12* pourrait être analogue à celles des gènes homéotiques de *Drosophila*.

Le gène *lin-12* comporte un ensemble de 11 répétitions de type EGF dont le consensus est montré dans la Figure 34.13. Il est en fait plus proche du consensus *Notch* que de tout autre membre de la famille de type EGF. Comme avec le gène *Notch*, les répétitions de type EGF ne constituent qu'une partie de la protéine.

Quelle est la signification de la présence de séquences ressemblant à des facteurs de croissance dans des gènes homéotiques d'insectes et de nématodes ? Nous avons eu tendance à considérer que les gènes homéotiques codaient pour des protéines contrôlant l'activité d'autres gènes. Si les répétitions de type EGF constituent le composant régulateur significatif de *Notch* et de *lin-12*, ces gènes pourraient bien fonctionner d'une manière différente.

Une possibilité est que ces protéines contrôlent le destin des cellules. Par exemple, si les peptides de type EGF subissent le même sort qu'EGF lui-même et sont clivés à partir du polypeptide répété, ils peuvent être nécessaires pour stimuler, après diffusion vers certaines cellules, le développement d'un organe. En leur absence ces cellules pourraient mourir ou, peut-être, elles ou leurs voisines pourraient se développer différemment.

Cette possibilité attire l'attention sur une question importante : ces produits de gènes de régulation fonctionnent-ils dans les cellules dans lesquelles ils ont été synthétisés ? Les produits du locus *BX-C* fonctionnent en effet *in situ* ainsi qu'en témoigne la localisation des transcrits et des protéines aux cellules des compartiments dont ils contrôlent le développement. (Formellement ces produits sont dits exercer un contrôle cellulaire autonome.) Il y a des arguments indiquant que *Notch* autonome.) Il y a des arguments indiquant que *Notch* pourrait fonctionner de la même façon et non selon un modèle de diffusion. Nous ne savons pas à quel endroit du nématode est exprimé le gène *lin-12* mais ceci devrait être aisé à déterminer en utilisant comme sonde une séquence exonique.

Si les séquences de type EGF demeurent intactes en tant que région de la protéine entière, il est probable, au moins dans le cas de *Notch*, qu'elles sont localisées à l'extérieur de la cellule. Une région transmembranaire, est qui pourrait ancrer la protéine à la surface cellulaire, est localisée plus loin dans la molécule. Un tel modèle soulève la question de l'action de cette protéine comme

un récepteur pour un quelconque facteur diffusible, ce qui serait en accord avec la régulation cellulaire autonome.

En fonction des homologies de séquences nous pouvons identifier deux classes nettement différentes de protéines de régulation :

- *Notch* et *lin-12* font partie d'un groupe dont les fonctions (qu'elles soient ou non de même nature qu'un facteur de croissance) sont probablement exercées à la face externe de la cellule, soit sur un mode cellulaire autonome, soit par diffusion.
- Le locus *BX-C* et certains gènes de segmentation de *Drosophila* codent pour des protéines qui sont localisées dans le noyau et qui sont généralement considérées comme de probables protéines se liant à l'ADN et exerçant des fonctions de régulation de l'expression des gènes.

UN MOTIF CODANT COMMUN : LA BOÎTE HOME0 (HOME0 BOX)

Une relation potentielle entre les fonctions de certains gènes de régulation s'est fait jour avec la découverte d'homologies de séquences dans les loci homéotiques et de segmentation. Une région de 180 pb, localisée près de l'extrémité 3' de l'unité de transcription de plusieurs loci, est commune à tous les membres de la famille. Cette **boîte homeo (homeo box)** fait partie d'une séquence ouverte de lecture dans laquelle beaucoup des changements de bases entre les différents gènes sont localisés au niveau de la troisième base du codon.

Il y a environ 10 gènes qui contiennent des homeo box chez *D. melanogaster*. Ils comprennent trois gènes dans le complexe *BX-C* (*Ubx*, *iab2* et *iab7*), trois dans ou près du complexe *ANT-C* (*Antp*, *Dfd*, et *Scr*) et d'autres encore dans le groupe formé de *ftz*, *engrailed* et *relié à engrailed*.

Il existe au moins deux groupes d'homeo box. Toutes celles dont nous avons parlé ci-dessus (à l'exception d'*engrailed* et de *relié à engrailed*) sont très homologues ; elles comprennent aussi le groupe *Antennapedia*. Dans des comparaisons par paires de séquences correspondantes de loci individuels, l'homologie est supérieure à 70-80 %, tant au niveau des acides aminés que des nucléotides. Les homeo box de *engrailed* et de *relié à engrailed* se ressemblent l'une l'autre mais sont en revanche plus éloignées de celles de l'autre groupe.

Les homeo box ont des tailles constantes et sont caractérisées par leur séquence basique. La Figure 34.14 illustre la conservation dans le groupe *Antennapedia*. Les acides aminés basiques forment environ 30 % du total,

Glu	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu
*		*																	
+	+		+																
Phe	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Ile	Glu	Ile	Ala	His	Ala	Leu	Cys
		*											*						
						+							+			+			+
Leu	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Ser	Lys	Lys	Glu
*					*											*			
+																+			+

Figure 34.14

L'homeo box est une séquence bien conservée de 60 acides aminés comprise dans plusieurs gènes homéotiques et de segmentation de *D. melanogaster* ainsi que dans des gènes eucaryotes de fonction inconnue. La séquence en acides aminés de l'homeo box *Antp* est donnée en entier ; les positions où ont lieu des substitutions sont indiquées pour *Ubx* (*) et pour *ftz* (+). Presque toutes les substitutions sont conservatives.

ce qui a été pris comme argument de l'aptitude de ce domaine à se lier à l'ADN. Cette idée est également conforme au sentiment que les homeo box résident dans des gènes aux fonctions similaires et dont les produits pourraient être de reconnaître un même motif d'ADN. Une autre possibilité est que le domaine homeo pourrait conférer la propriété de reconnaître une protéine ubiquitairement nécessaire à l'expression des gènes, par exemple l'ARN polymérase ou peut-être certains constituants structuraux de la chromatine.

Un remarquable développement quant à la signification des homeo box est constitué par la découverte qu'une sonde d'ADN spécifique de l'homeo box hybride avec de l'ADN de grenouille, de souris et d'homme. Certains des gènes d'amphibiens et de mammifères qui contiennent des homeo box ont été clonés. Les homeo box humaines et murines sont localisées dans des groupes gènes. Une région de 20 kb d'ADN humain contient trois homeo box et une région de 3 kb d'ADN murin en contient 2 à 5. La séquence des homeo box de mammifères est à environ 70 % homologue de celle de *D. melanogaster*. Nous ne savons pas encore exactement si ces homeo box groupées représentent de multiples copies localisées dans une seule protéine ou bien identifient des groupes de gènes dont chacun contient une seule homeo box.

Les groupes humains et murins n'ont pas d'homologie avec les séquences de *D. melanogaster* à l'exception des homeo box. Comparés l'un à l'autre cependant, ils portent d'autres régions d'homologie et peuvent ainsi représenter des fonctions correspondantes chez différents mammifères. Les transcrits de ces régions apparaissent précocement au cours du développement embryonnaire, ce qui suggère qu'ils peuvent avoir des fonctions grossièrement analogues à celles des gènes de la mouche

c'est-à-dire qu'ils pourraient être eux aussi impliqués dans l'embryogenèse.

Une observation intéressante est représentée par une autre homeo box chez la souris (dont le gène *Mo-en.1*) qui n'est homologue à aucune autre séquence portant des homeo box. C'est avec l'homeo box de *engrailed* qu'elle présente le plus de ressemblance, les 20 acides aminés du côté C-terminal de l'homeo box étant homologues dans les deux gènes. Il se pourrait par conséquent que ce gène murin fût en quelque sorte un équivalent du gène *engrailed* de la mouche.

Deux gènes de *X. laevis* ont été clonés grâce à l'homeo box qu'ils contenaient. Ils possèdent de 55 à 59 résidus de la séquence d'acides aminés de l'homeo box *Antp*. Ces deux types de gènes sont exprimés dans l'ovocyte ou précocement au cours du développement embryonnaire.

Quelles fonctions communes pourraient avoir les gènes intervenant dans le contrôle du développement des mouches du vinaigre, des amphibiens et des mammifères ? Peut-être les produits de ces gènes donnent-ils un certain type d'instructions universelles indiquant à la cellule qu'elle doit arrêter de se diviser, se diviser plus rapidement ou geler un état donné du développement.

Les gènes homéotiques sont considérés comme des régulateurs des commutations se produisant au cours du développement, opérant dans des compartiments spécifiques et codant peut-être pour des protéines qui se lient à des batteries particulières de gènes dont les produits donnent leur identité particulière aux cellules de chaque compartiment. Des mécanismes de même type pourraient-ils opérer chez les vertébrés et les invertébrés en dépit de leur organisation très différente ? Avons-nous découvert là un mécanisme commun du contrôle du développement embryonnaire ?

POUR EN SAVOIR PLUS

Les notions fondamentales dans le développement de *Drosophila* ont été passées en revue par Mahowald & Hardy (*Ann. Rev. Genet.* **19**, 149-177, 1985). Le locus *white* a été séquencé par O'Hare et al. (*J. Mol. Biol.* **180**, 437-455, 1984) et l'insertion DZL a été caractérisée par Bingham & Zachar (*Cell* **40**, 819-825, 1985). Le locus *BX-C* a été cloné par Bender & al. (*Science*

221, 23-29, 1983) et caractérisé par Beachy & al. (*Nature* **313**, 545-551, 1985) et Karch & al. (*Cell* **43**, 81-96, 1985). *ANT-C* a été caractérisé par Scott & al. (*Cell* **35**, 763-766, 1983). Les homeo box de *Drosophila* et les gènes qui les contiennent ont été expliqués par Gehring (*Cell* **40**, 3-5, 1985). Les homeo box de vertébrés ont été passées en revue par Manley & Levine (*Cell* **43**, 1-2, 1985).

CHAPITRE 35

ONCOGÈNES : EXPRESSION ANORMALE DES GÈNES ET CANCER

Une des caractéristiques majeures des eucaryotes supérieurs est que leur temps de vie est défini ; cette propriété est partagée par toutes les cellules somatiques, cellules dont la croissance et la division sont hautement régulées. Seules les cellules cancéreuses, qui ne sont autres que des variants ayant perdu le contrôle normal de leur croissance, font exception. Celles-ci sont capables de se développer dans des endroits inhabituels et de se propager indéfiniment, deux propriétés pouvant être létales pour l'organisme chez lequel elles apparaissent.

Afin de caractériser les événements qui permettent à la cellule d'échapper à un contrôle normal de sa croissance et engendrent des tumeurs, il est nécessaire d'avoir un système permettant *in vitro* la croissance des cellules cancéreuses. A l'heure actuelle, nous avons le choix entre deux systèmes, ni l'un ni l'autre ne donnant entière satisfaction :

- « Les cellules primaires » — qui descendent directement de cellules prélevées à partir de l'organisme — peuvent reproduire fidèlement le phénotype *in vivo*, mais ne survivent bien souvent que très peu de temps.
- Les cellules qui se sont « établies » et forment une lignée cellulaire (non tumorigène) peuvent se perpétuer indéfiniment, mais leurs propriétés peuvent se modifier lors de l'adaptation des cellules à la culture.

Une lignée cellulaire non tumorigène établie subissant un contrôle de croissance montre des signes caractéristiques dont font souvent partie :

La dépendance à l'ancrage — une surface solide ou ferme est nécessaire aux cellules pour qu'elles s'y attachent.

* N. d. T. : Des milieux chimiquement définis peuvent, en présence des hormones et des facteurs de croissance nécessaires, permettre la croissance et la prolifération cellulaire en absence de sérum.

La dépendance au sérum — le sérum est nécessaire à la croissance cellulaire car il apporte les facteurs de croissance essentiels à la cellule *.

La saturation à une certaine densité — les cellules arrêtent de se diviser à une certaine densité, car leur croissance est inhibée par des mécanismes impliquant peut-être les contacts inter-cellulaires.

Ces propriétés sont autant de paramètres grâce auxquels la normalité d'une cellule pourra être estimée. Bien sûr, toute lignée cellulaire établie ne peut fournir qu'une estimation des contrôles qui s'effectuent *in vivo*. La prudence qu'il faut observer lors de l'analyse des bases génétiques du contrôle de croissance dans de telles lignées est soulignée par le fait que ces cellules ne sont pas de réels diploïdes, en effet les chromosomes homologues synthétisés lors de la mitose subissent presque toujours des modifications.

Certaines, voire toutes les propriétés précédemment citées, sont modifiées dans les cellules mises en culture à partir de tumeurs et non de tissus normaux. Ces cellules sont dites **transformées**. Une cellule transformée pousse dans des conditions moins limitantes qu'une cellule normale : elle se divise plus souvent, ne nécessite pas toujours une surface solide à laquelle s'attacher, peut avoir une dépendance au sérum réduite, peut se diviser en s'amoncelant et ainsi former une masse cellulaire solide au lieu d'être réduite à pousser en couche mince ou en monocouche sur une surface. Sur la **Figure 35.1**, sont comparés un fibroblaste normal et un variant « transformé » en culture.

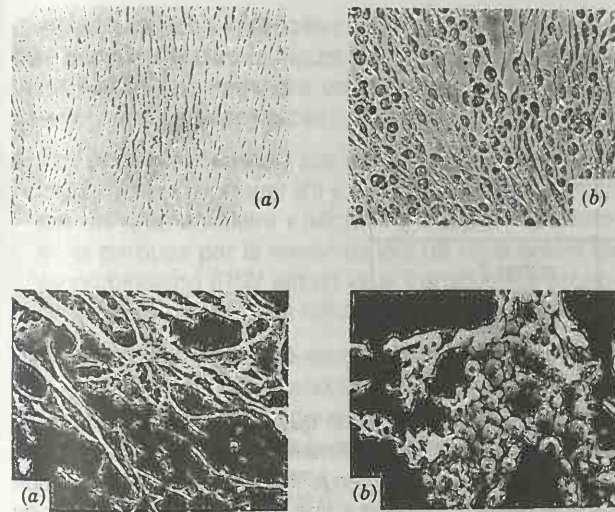


Figure 35.1

Les fibroblastes normaux poussent en une couche plate de cellules disséminées alors que les fibroblastes transformés se rassemblent et poussent en masses cellulaires.

Les cultures figurant sur la gauche sont constituées de cellules normales, celles de droite de cellules transformées. Les photographies du haut ont été prises en microscopie optique, celles du bas en microscopie électronique à balayage.

Photographies aimablement fournies par Hidesaburo Hanafusa et J. Michael Bishop.

Certains traitements convertissent des cellules normales en cellules transformées, ces dernières deviennent alors des modèles pour l'étude des mécanismes impliqués dans la cancérisation. Il est possible de classer les modifications qui apparaissent en deux catégories : celles qui concernent l'**immortalisation**, et celles qui concernent la **transformation**. L'immortalisation est la propriété de croissance indéfinie ; elle est acquise par toute lignée cellulaire établie, qui ne nécessite alors que la transformation pour être tumorigène. Une cellule primaire doit passer par ces deux étapes.

Les cellules en culture ayant subi des modifications dues à l'immortalisation et à la transformation représentent des modèles pour l'étude des tumeurs animales. En comparant des lignées de cellules transformées et des cellules normales, nous espérons mettre en évidence les bases génétiques de la formation des tumeurs, et également comprendre les mécanismes phénotypiques impliqués dans la conversion d'une cellule normale en une cellule cancéreuse.

Il serait simpliste de supposer que la base de la formation des cellules cancéreuses est uniforme ; diverses modifications dans la constitution de la cellule peuvent être responsables de la formation de tumeur. Cependant, certaines caractéristiques sont communes à toutes

les cellules cancéreuses. Une caractéristique majeure du changement qui s'opère quand une cellule se transforme en cellule cancéreuse est qu'elle et sa descendance n'observent plus les contraintes normales de croissance. D'autre part, elle peut sortir du tissu normal et **métastaser**, c'est-à-dire partir du tissu d'origine et établir des colonies à d'autres endroits de l'organisme.

Quelle est la cause de ce premier changement du phénotype cellulaire ? Divers agents sont carcinogènes ; produisent-ils des modifications épigénétiques ou agissent-ils, directement ou indirectement, pour changer le génotype de la cellule ? Les carcinogènes sont parfois divisés en deux catégories : les initiateurs et les promoteurs de la formation tumorale. De multiples changements sont souvent nécessaires dans la création d'un cancer ; et quelquefois, les virulences accrues des tumeurs sont le résultat de la somme de plusieurs changements.

Jusqu'à récemment, nous n'avions aucune information quant aux bases moléculaires du cancer. Nous connaissons maintenant l'existence d'une classe de gènes dans lesquels des mutations peuvent être la cause de la transformation. Ces gènes furent identifiés grâce à l'analyse de virus transformants, car il fut rapidement mis en évidence que certains d'entre eux avaient des équivalents cellulaires pouvant être impliqués dans la formation de tumeurs.

LES VIRUS TRANSFORMANTS PEUVENT ÊTRE PORTEURS D'ONCOGÈNES

La transformation peut se produire spontanément, être induite par certains agents chimiques et, plus fréquemment, résulter d'une infection par un virus tumorigène. Il existe de nombreuses classes de virus tumorigènes, incluant des virus à ADN et des virus à ARN, présents chez les aviaires et plus généralement dans le règne animal.

Qu'est-ce qui permet au virus tumorigène de transformer sa cellule cible ? Le virus n'est pas porteur d'un ensemble de gènes codant pour des produits directement responsables des nombreuses modifications nécessaires au changement du phénotype cellulaire. Le virus initie une série d'événements qui est ensuite exécutée par des protéines cellulaires. Schématiquement, nous supposons que le virus oriente vers une certaine direction la régulation cellulaire, ce qui va avoir pour effet de changer les propriétés de croissance de la cellule cible.

L'activité transformante du virus tumorigène est peut-être le fait d'un ou de gène(s) particulier(s) porté(s) par le génome viral. Les gènes qui confèrent la capacité à

Tableau 35.1

Les virus transformants peuvent être porteurs d'oncogènes.

Classe virale	Type de virus	Taille du génome	Oncogènes
Adéno	ADN double brin	~ 37 kb	E1A & E1B
Papova	ADN double brin	5-6 kb	antigènes T
Herpès	ADN double brin	~ 160 kb	aucun ?
Rétro	ARN simple brin	6-9 kb	spécifique à chacun

convertir des cellules normales en cellules tumorales sont appelés des **oncogènes**. Sur le **Tableau 35.1** sont résumées les propriétés générales de quatre classes majeures de virus transformants.

Les adénovirus ont deux types d'activité, E1A et E1B, qui sont ensemble responsables de la transformation. Les papovavirus synthétisent des protéines appelées antigènes, dont les séquences codantes se chevauchent et qui partagent également la responsabilité dans la transformation. Aucune activité transformante n'a été décelée en ce qui concerne les virus herpès*. Les fonctions transformantes de tous ces virus à ADN n'ont dans aucun cas pu être apparentées à une fonction cellulaire.

La situation est différente en ce qui concerne les rétrovirus. Ils sont divisés en différentes catégories selon le type de tumeur qu'ils provoquent chez l'animal : leucémie, sarcome, carcinome, etc., ce qui est en accord avec le fait que chaque virus est spécifique d'un type particulier de cellule. En général, un rétrovirus n'est constitué que d'un seul oncogène issu d'un gène cellulaire et capturé par le rétrovirus lors d'un événement de transduction.

Les rétrovirus peuvent transmettre une information génétique horizontalement et verticalement, comme l'illustre la **Figure 35.2**. La transmission horizontale est effectuée grâce au processus normal d'infection virale. La transmission verticale se produit lorsqu'un virus s'intègre dans la lignée germinale d'un organisme sous la forme d'un provirus ; tel un bactériophage lysogénique, il est alors hérité par la descendance comme un locus mendélien (voir Chapitre 30).

Le cycle de vie rétroviral permet la propagation d'une information génétique à travers l'ARN et l'ADN. Une infection rétrovirale se déroule en différentes étapes illustrées sur la Figure 30.7 ; l'ARN subit une transcription

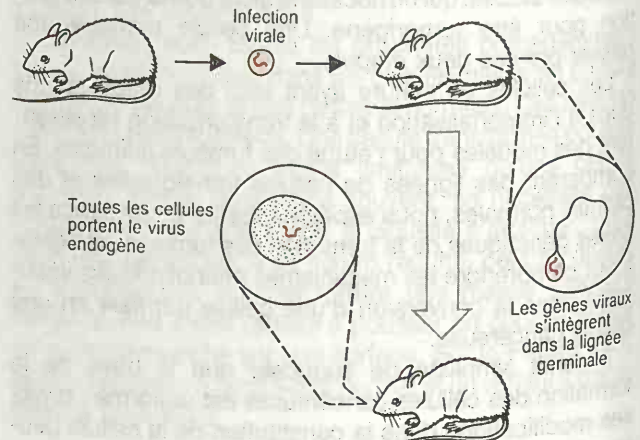
inverse en ADN simple brin qui sera ensuite converti en ADN double brin et finalement intégré dans le génome où il pourra être transcrit en ARN infectueux. L'intégration dans le génome conduit à une transmission verticale du provirus ; l'expression du provirus peut engendrer des particules rétrovirales actives qui seront transmises horizontalement.

La propagation de l'information génétique grâce à un intermédiaire ARN n'est pas limitée uniquement aux séquences rétrovirales actives. Certains transposons (partageant sûrement des origines communes avec les rétrovirus) suivent le même schéma général, cependant leurs mouvements semblent limités au génome d'une seule cellule et il ne paraît pas y avoir d'étape d'infection. Le comportement de l'élément Ty chez la levure est l'exemple le mieux caractérisé (voir Chapitre 30).

Durant un cycle infectueux, un rétrovirus peut échanger

Figure 35.2

Les rétrovirus transmettent horizontalement une information génétique en infectant de nouveaux hôtes ; l'information est transmise verticalement si un virus s'intègre dans le génome de la lignée germinale.



* N. d. T. : Cette affirmation n'est plus valable en 1988.

une partie de ses séquences contre une séquence cellulaire comme l'illustre la Figure 30.12. Ce genre d'événement est rare, il engendre un virus transducteur qui a deux propriétés importantes :

- Il ne peut généralement pas se répliquer par lui-même, mais peut se propager s'il y a simultanément infection par un virus auxiliaire « helper » qui apporte les fonctions perdues par le rétrovirus lors de l'événement de recombinaison (RSV est un virus transducteur exceptionnel qui a gardé ses fonctions de répllication).
- Lors d'une infection, le virus transducteur est porteur des gènes cellulaires acquis lors de l'événement de recombinaison, gènes dont l'expression peut modifier le phénotype de la cellule infectée.

Si l'information génétique du virus transducteur stimule la croissance des cellules cibles, cela peut constituer un avantage sélectif pour le virus au cours des cycles infectueux ultérieurs. Si un virus possède un gène dont le produit est impliqué dans la stimulation de la croissance d'un type particulier de cellule, ce gène, dans sa forme capturée ou après introduction de mutations, peut rendre

le virus capable de se propager en stimulant la croissance des cellules infectées.

Bien sûr, la transformation n'est pas le seul mécanisme selon lequel les rétrovirus sont capables d'affecter leurs hôtes. Un exemple notable est le virus du SIDA (HIV1 et HIV2) qui appartient à la classe des lentivirus. Le virus du SIDA infecte et tue les lymphocytes T portant le récepteur T4, il détruit le système immunitaire de l'hôte. L'ADN du virus est constitué des régions habituelles *gag-pol-env*, mais il comporte également deux autres cadres de lectures, auxquels sont attribuées les fonctions de létalité *.

LES ONCOGÈNES RÉTROVIRAUX ONT DES ÉQUIVALENTS CELLULAIRES

Le gène codant pour l'activité rétrovirale responsable de la transformation des cellules cibles par le rétrovirus peut être identifié par analyse mutationnelle ; ainsi on a pu déterminer chez certains rétrovirus que le gène responsable de l'activité transformante était un oncogène ; les oncogènes (ou supposés oncogènes) portés par certains rétrovirus sont résumés sur le **Tableau 35.2**.

Tableau 35.2

Tous les rétrovirus transformants sont porteurs d'un oncogène dérivé d'un gène cellulaire.

Virus	Nom	Espèce	Tumeur	Oncogène
Sarcome de Rous	RSV	poulet	sarcome	<i>src</i>
Sarcome murin de Harvey	Ha-MuSV	rat	sarcome et leucémie	<i>H-ras</i>
" de Kirsten	Ki-MuSV	rat	érythrocytaire	<i>K-ras</i>
" de Moloney	Mo-MuSV	souris	sarcome	<i>mos</i>
Ostéosarcome murin FBJ	FBJ-MuSV	souris	chondrosarcome	<i>fos</i>
Sarcome simien	SSV	singe	sarcome	<i>sis</i>
Sarcome félin	PI-FeSV	chat	sarcome	"
"	SM-FeSV	chat	fibrosarcome	<i>fms</i>
"	ST-FeSV	chat	"	<i>fes</i>
Sarcome de Fujinami	FuSV	poulet	sarcome	<i>fps</i>
Myélocytomatose aviaire	MC29	poulet	carcinome, sarcome, et myélocytome	<i>myc</i>
Leucémie d'Abelson	MuLV	souris	lymphome des cellules B	<i>abl</i>
Réticuloendothéliose	REV-T	dinde	leucémie lymphatique	<i>rel</i>
Erythroblastose aviaire	AEV	poulet	leucémie érythrocytaire et fibrosarcome	<i>erbB</i> (<i>erbA?</i>)
Myéloblastose aviaire	AMV	poulet	leucémie myéloblastique	<i>myb</i>

Les noms et abréviations des virus sont le reflet de l'histoire de leur isolement et du type de tumeur dont ils sont la cause. Tous ces rétrovirus à l'exception de AEV ont un seul oncogène.

* N. d. T. : Les virus HIV possèdent en réalité un beaucoup plus grand nombre de séquences ouvertes codant pour des protéines dont certaines sont indispensables au cycle rétroviral. Par exemple, le gène *tat* code pour une protéine transactivatrice stimulant la transcription du génome rétroviral.

Le(s) gène(s) cellulaire(s) capturé(s) par un virus transducteur peu(ven)t être identifié(s) en comparant sa (leur) séquence avec celle du virus parental. Il y a invariablement une nouvelle région qui est fortement apparentée à des séquences du génome cellulaire. Les séquences cellulaires ne sont pas oncogéniques — si elles l'étaient, l'organisme n'aurait jamais pu survivre — elles doivent être considérées comme des **proto-oncogènes**, c'est-à-dire comme des gènes dont la capture par un rétrovirus et les modifications qui en résultent peuvent les transformer en oncogènes.

Les oncogènes viraux et leurs équivalents cellulaires sont décrits en utilisant les préfixes suivants : *v* pour viral et *c* pour cellulaire. Par exemple, l'oncogène porté par le virus du sarcome de Rous est appelé *v-src*, et le gène qui lui est apparenté dans les génomes cellulaires *c-src*.

Comme il est décrit dans le Chapitre 30, le gène *v-onc* est capturé sous la forme d'un ARN lors d'une infection rétrovirale. Si l'on compare une séquence *v-onc* avec sa séquence *c-onc* correspondante, il est inévitablement mis en évidence que l'organisation du gène viral correspond à celle de l'ARNm plutôt qu'à celle du gène *c-onc*. Les gènes *v-onc* consistent donc en des séquences codantes ininterrompues alors que les gènes *c-onc* sont organisés selon le modèle habituel exon-intron.

L'expression d'une séquence *v-onc* donne souvent lieu à une protéine de fusion portant également des fonctions virales. La structure la plus couramment observée est une fusion *gag-v-onc* (voir Tableau 35.9). Lorsque la séquence codante *v-onc* est étroitement apparentée à la séquence d'un gène ou d'une partie d'un gène du génome cellulaire de l'hôte, il peut être possible de définir les sites où se sont produits les événements de recombinaisons qui ont permis l'insertion de la séquence cellulaire dans le virus.

Dans les cellules normales, les gènes *c-onc* sont généralement exprimés à des taux relativement bas ; la quantité de protéine synthétisée par un gène *v-onc* est souvent plus importante, rendant compte du taux d'expression du génome rétroviral lors de l'infection.

Plus de 20 gènes *c-onc* ont été identifiés jusqu'à maintenant grâce à leur présence dans des rétrovirus. Quelquefois, le même gène *c-onc* est présent dans différents virus transformants, par exemple, le virus de singe SSV et la lignée PI du virus félin FeSV portent tous deux un *v-onc* issu de *c-sis*. Souvent, plusieurs lignées

d'un virus transformant sont isolées et chacune porte un gène *v-onc* dérivé de *c-onc* différents.

Parfois, les gènes *v-onc* de différents virus peuvent être apparentés comme c'est le cas pour les lignées Harvey et Kirsten de MuSV qui portent les gènes *v-ras* dérivés de deux membres différents de la famille des gènes cellulaires *c-ras*. Dans d'autres cas, les gènes *v-onc* de virus apparentés correspondent à des ancêtres cellulaires non apparentés ; par exemple, trois lignées différentes de FeSV dérivent probablement du même virus mais ont transduit les oncogènes *sis*, *fms* et *fes*. Les événements impliqués dans la formation d'un virus transducteur peuvent être complexes ; certains virus ont des séquences qui dérivent de plusieurs gènes cellulaires*.

Si l'on considère la rareté d'un événement de transduction, le fait de trouver dans différents isolats le même gène *c-onc* est significatif. Par exemple, certains virus portent des gènes *v-myc* dont les extrémités sont différentes mais dont la partie commune ne diffère que par un petit nombre de mutations ponctuelles. L'existence de tels isolats est la preuve que tous les gènes de type *c-onc* ont probablement été identifiés**.

Quel est le degré de parenté entre un gène *v-onc* et le gène *c-onc* correspondant ? Dans plusieurs cas, la

Tableau 35.3

Les régions codantes des *v-onc* ont une homologie importante avec les gènes *c-onc* qui leur correspondent.

Gène	Nbre de codons dans <i>c-onc</i>	Nbre de codons dans <i>v-onc</i>	Nbre d'acides aminés conservés	Homologie
<i>mos</i>	369	369	358	97 %
<i>H-ras</i>	189	189	186	98 %
<i>K-ras</i>	189	189	182	96 %
<i>sis</i>	220	220	202	92 %
<i>src</i>	533	514	498	97 %
<i>erbB</i>	1210	600	501	83 %
<i>myc</i>	417	417	415	99 %
<i>myb</i>	~ 1000 ?	372	361	97 %

Les codons dans *c-onc* correspondent à la longueur de toute la région codante du gène cellulaire. Le nombre de codons dans *v-onc* correspond à la longueur de la région de la protéine virale homologue à *c-onc*. Le nombre d'acides aminés conservés et le pourcentage d'homologie sont le reflet du nombre de positions identiques entre les protéines.

* N. d. T. : Certains rétrovirus oncogéniques possèdent plusieurs oncogènes ; par exemple, le rétrovirus AEV (virus de l'érythroblastose aviaire) possède les oncogènes *v-erbB* et *v-erbA*, tous deux transduits à partir des séquences cellulaires *c-erbB* et *c-erbA* dont nous reparlerons.

** N. d. T. : La découverte entre 1986 et 1988 de quelques nouveaux oncogènes montre que l'affirmation que tous les oncogènes ont été identifiés doit être prise avec prudence.

séquence codante de *v-onc* semble avoir relativement peu évoluée par rapport à la séquence *c-onc*, évolution qui se manifeste généralement par des mutations ponctuelles. Le fait que la majorité des modifications de nucléotides soient situées à la position de la troisième base du codon où elles n'affectent pas la séquence protéique, indique que les gènes *v-onc* et *c-onc* partagent probablement la même fonction. Sur le **Tableau 35.3** sont résumées les relations qui existent entre les gènes *v-onc* et *c-onc* dans les cas où tous deux ont été séquencés.

Aux gènes *mos*, *ras*, *sis* et *myc* correspondent des cas où le gène *c-onc* a été entièrement capturé par le virus et où celui-ci n'a subi que peu de modifications affectant son produit. Les gènes *src* viral et cellulaire sont également de même longueur mais les 19 acides aminés C-terminaux de *c-src* ont été remplacés par une autre séquence de 12 acides aminés dans *v-src*.

ErbB illustre l'exemple où seule une partie du gène *c-onc* a été capturée par le virus. Le gène *v-onc* correspond à la moitié C-terminale de son équivalent cellulaire. Nous nous attendons, par conséquent, à ce que son produit n'ait qu'une partie des capacités fonctionnelles du produit de *c-onc*. Un autre exemple de gène *v-onc* tronqué est celui de *v-myc* qui a perdu les séquences N-terminale et C-terminale du gène *c-myc*. Nous pouvons proposer deux modèles afin d'expliquer la différence de propriétés entre les gènes *v-onc* et *c-onc* :

- Un modèle quantitatif qui suppose que les gènes viraux sont indiscernables des gènes cellulaires et que leur oncogénicité vient du fait qu'ils s'expriment en quantité beaucoup plus importante (ce qui est effectivement souvent le cas) ou dans des types cellulaires inappropriés.
- Un modèle qualitatif qui suppose que les gènes *c-onc* n'ont pas intrinsèquement de propriétés oncogéniques mais peuvent être convertis par mutation en oncogènes dont les effets dévastateurs démontrent l'acquisition de nouvelles propriétés. Bien sûr, ces modèles ne sont pas mutuellement exclusifs mais chacun d'entre eux peut s'appliquer seul pour certains oncogènes.

Qu'est-ce qui permet de dire que la séquence *v-onc* est responsable du potentiel oncogénique d'un virus transformant ? Afin de déterminer ce qui confère l'oncogénicité, deux stratégies ont été adoptées :

- Muter la séquence *v-onc* afin d'éliminer l'activité transformante.
- Trouver les conditions dans lesquelles le gène *c-onc* devient oncogénique en modifiant, par exemple, la séquence *c-onc* (souvent par échange de séquence avec *v-onc*).

La preuve directe que l'expression de la séquence *v-onc* est responsable de la transformation fut apportée par RSV. Des mutations thermo-sensibles dans *v-src* permettent au phénotype transformé de disparaître si l'on augmente la température ou de réapparaître si on la diminue. Cela montre clairement que le gène *v-src* est nécessaire pour initier mais aussi maintenir l'état transformé.

Les modifications de la séquence *v-src* sont nécessaires à son oncogénicité. L'expression de *c-src* à des taux dix fois plus importants que la normale ne transforme pas les cellules alors que *v-src* est oncogénique à des niveaux d'expression beaucoup plus bas.

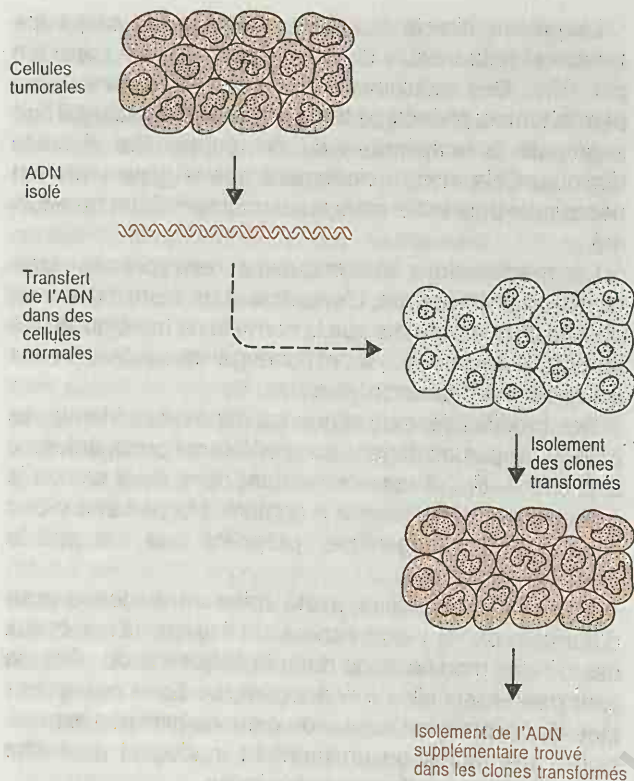
Des modifications de séquence dans *c-ras*, identiques à celles apparues dans *v-ras*, confèrent l'oncogénicité à la protéine. *Src* et *ras* représentent donc deux cas où le changement de séquence a conféré à la protéine *v-onc* sa propriété oncogénique, propriété que n'a pas la protéine *c-onc*.

Une relation similaire a été mise en évidence pour d'autres *c-onc* et *v-onc*, cependant il existe de nombreux cas où des modifications dans la séquence du gène ne sont pas essentielles à l'oncogénicité. Dans ces exemples-là, l'expression accrue du gène ou bien son expression à un endroit ou un moment inadéquat peut être responsable des effets oncogéniques.

LES PROTO-ONCOGÈNES RAS PEUVENT ÊTRE ACTIVÉS PAR MUTATION

Les oncogènes peuvent-ils être responsables des événements de transformation autrement que par infection virale ? Il est possible de tester directement si des gènes sont transformants en les transfectant dans des cellules receveuses « normales » (en réalité, on utilise la lignée établie de fibroblastes de souris 3T3). Le protocole de cette technique est illustré de façon schématique sur la **Figure 35.3**. Si l'expression d'un gène spécifique engendre un état transformé, son introduction dans de nouvelles cellules devrait les transformer.

Quand une cellule est transformée dans une culture 3T3 (ou quelque autre culture « normale »), ses descendants s'amassent pour former un **foyer**. L'apparition de foyers permet de mesurer la capacité transformante d'une préparation d'ADN. Si l'on part d'une préparation d'ADN isolé à partir de cellules tumorales, l'efficacité de formation de foyers est faible. Toutefois, si l'on isole et l'on clone le gène transformant, de meilleures efficacités peuvent être obtenues. En réalité, le « pouvoir » transformant d'un gène peut être caractérisé par l'efficacité de

**Figure 35.3**

Les oncogènes peuvent être isolés directement en testant l'ADN des cellules tumorales quant à leur capacité à transformer des cellules normales en cellules tumorales.

formation de foyer de la séquence clonée. Un gène hautement transformant peut avoir une efficacité supérieure à 100 foyers/ng d'ADN/ 10^6 cellules, alors que l'efficacité d'un gène faiblement transformant peut être inférieure à 10 foyers/ng d'ADN/ 10^6 cellules.

Les propriétés de certains gènes transformants isolés par cette technique sont résumées sur le **Tableau 35.4**. L'ADN ayant une activité transformante peut être isolé

uniquement à partir de cellules tumorales ; il n'est pas présent dans les cellules normales.

Les gènes transformants ont deux propriétés connues :

- Ils sont étroitement apparentés à des séquences d'ADN trouvées dans les cellules normales. Cela signifie que le pouvoir transformant est dû à la mutation du gène normal (un proto-oncogène) qui le transforme en oncogène. La mutation peut être une mutation ponctuelle ou une réorganisation plus importante de l'ADN autour du gène *c-onc* comme le résume le Tableau.
- Ils peuvent avoir des équivalents dans les oncogènes portés par des virus transformants connus. Cela signifie que le répertoire des proto-oncogènes est limité, et que ce sont toujours les mêmes gènes dans le génome cellulaire qui sont la cible de mutations pouvant les transformer en oncogènes ou qui peuvent devenir des oncogènes viraux. La majorité des proto-oncogènes ont probablement été transduits au moins une fois par un rétrovirus. Trois des cinq exemples illustrés sur le Tableau ont des équivalents directs dans les gènes *v-onc* ; d'autre part, *N-ras* est membre d'une famille représentée dans les gènes *v-onc*.

Une grande proportion des oncogènes détectés par le test de transfection est dérivée de la famille *c-ras*, qui est constituée de trois membres actifs chez l'homme et le rat, dispersés dans le génome (il existe également des pseudogènes). La structure de tous ces gènes est très similaire, comme l'illustre la **Figure 35.4**. Les gènes N et H ont trois exons homologues ; le gène K a un exon supplémentaire, ce qui lui permet de coder, grâce à un épissage alternatif, pour des protéines se chevauchant. Les protéines connues sont très ressemblantes et ont toutes une taille d'environ 21 000 daltons. La majeure partie de la séquence est conservée entre les protéines ras, cependant il existe des régions variables.

Bien que membres d'une même famille, les gènes H, K, et N *c-ras* ont évolué différemment les uns par rapport

Tableau 35.4

Certains oncogènes peuvent être détectés grâce à leur capacité à transformer des fibroblastes de souris 3T3 dans un test de transfection.

<i>c-onc</i>	<i>v-onc</i>	Type de modif. dans <i>c-onc</i>	Type de tumeur portant l'oncogène
<i>H-ras</i>	<i>H-ras</i>	mutation ponctuelle	tumeurs humaines variées : voir Tableau 35.5 neuroblastome de rat plasmocytome murin ostéosarcome humain
<i>K-ras</i>	<i>K-ras</i>	mutation ponctuelle	
<i>N-ras</i>	aucun	mutation ponctuelle	
<i>neu</i>	aucun	mutation ponctuelle	
<i>mos</i>	<i>mos</i>	réarrangement	
<i>met</i>	aucun	réarrangement	

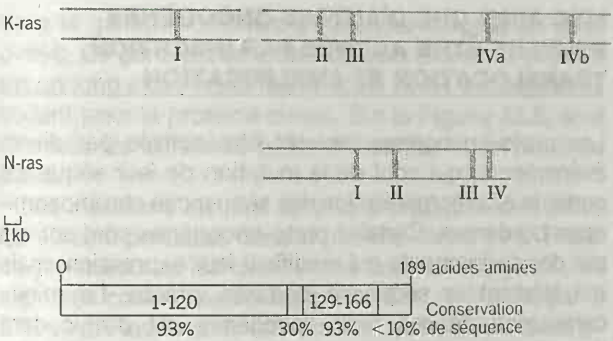


Figure 35.4
La famille des gènes *ras* humains est constituée de trois membres dont la structure est étroitement apparentée. Le gène *N-ras* a 4 exons mais le gène *K-ras* peut avoir un exon supplémentaire (IVb). Les trois protéines *ras* ne diffèrent que par des régions relativement restreintes.

aux autres. Les gènes H et K ont des équivalents *v-ras* portés respectivement par les lignées Harvey et Kirsten du virus de sarcome murin (voir Tableau 35.2). Chaque gène *v-ras* est étroitement apparenté au gène *c-ras* correspondant, avec seulement 3 à 7 changements d'acides aminés selon le cas (voir le Tableau 35.3). Cela signifie que les lignées des virus Harvey et Kirsten sont issues d'événements de recombinaison indépendants grâce auxquels chaque virus ancestral a capturé l'une des séquences *c-ras*.

Les variants oncogéniques des gènes *c-ras* ont été isolés à partir de préparations d'ADN transformant issues de diverses lignées cellulaires tumorales. Chaque proto-oncogène *c-ras* peut donner lieu à un oncogène transformant par la mutation d'une seule base. Sur le **Tableau 35.5**, sont indiquées les différentes modifications d'acides aminés dues à des mutations provoquant l'oncogénicité ainsi que les acides aminés correspondant aux mêmes positions sur les gènes *v-H-ras* et *v-K-ras*. Le plus indique que l'acide aminé est le même que celui présent dans le *c-ras* de type sauvage.

Dans différentes tumeurs humaines, on trouve une mutation dans le gène *ras* qui engendre la substitution d'un acide aminé dans l'une des protéines *ras*. Ces changements oncogéniques apparaissent à deux positions, les positions 12 et 61.

Dans de multiples tumeurs chez le rat, l'on trouve des mutations situées aux mêmes positions sur les gènes *ras* mutés. Cela suggère que la protéine *ras* normale est potentiellement hautement convertible en une forme tumorigène par la mutation de l'un de ses codons chez le rat ou l'homme (et peut-être tout mammifère).

L'on définit la tumorigénicité des différentes formes de

Tableau 35.5
Les gènes de la famille *ras* peuvent devenir oncogènes dans des tumeurs humaines et sont portés par des rétrovirus transformants.

Proto-oncogène	Acide aminé			Origine tumorale
	12	59	61	
<i>c-ras</i> (H, K, N)	Gly	Ala	Gln	séquence de type sauvage
<i>H-ras</i>	+	+	Leu	carcinome pulmonaire Hs242
<i>K-ras</i>	Val	+	+	carcinome de la vessie T24
	Cys	+	+	carcinome pulmonaire Calu-1
	Arg	+	+	carcinome pulmonaire LC-10
	Val	+	+	carcinome du côlon SW480
<i>N-ras</i>	+	+	Lys	neuroblastome SK-N-SH
	+	+	Arg	carcinome pulmonaire SW1271
				virus de sarcome murin
<i>H-ras</i>	Arg	Thr	+	lignée Harvey
<i>K-ras</i>	Ser	Thr	+	lignée Kirsten

c-ras par leur capacité à induire une tumeur quand elles sont transfectées dans une nouvelle cellule. Le fait que différentes lignées cellulaires dérivant de tumeurs humaines et de rongeurs possèdent des gènes *c-ras* mutés permet de corréler la présence d'un oncogène actif à l'existence d'une tumeur. Cependant, cela ne prouve pas que l'oncogène est initialement responsable de la formation de la tumeur ; sa fonction peut simplement être l'une de celles qui doivent être activées pour former une tumeur et la propager.

Le cas suivant donne une meilleure indication de cause à effet : chez un patient ayant un carcinome pulmonaire, il a été mis en évidence dans l'ADN du tissu atteint la présence d'un gène *c-ras* muté, alors que dans l'ADN des tissus normaux le même gène était de type sauvage. La mutation du gène *c-ras* est apparue uniquement dans le tissu tumoral du patient et est certainement liée à l'apparition de la tumeur.

Si l'on compare les quelques positions qui sont modifiées dans les gènes *v-ras* de rat à celles des gènes *c-ras* humains, deux sites communs sont mis en évidence (12 et 59). La position 12 est la même que l'un des sites mutés dans les tumeurs humaines. Il semble donc que la mutation de cette position soit un facteur critique pour induire le pouvoir transformant des gènes *v-ras*.

Le principe général établi par ce travail est que la substitution d'un acide aminé peut convertir un proto-oncogène cellulaire en un oncogène. Si l'on considère les séquences de tous les variants transformants de *c-ras*, la position 12 des trois gènes *ras* représente un

site extrêmement puissant où une mutation peut transformer le gène en oncogène. Un tel oncogène peut être associé à l'apparition d'une tumeur spontanée dans l'organisme. Il peut également être porté par un rétrovirus et dans ce cas, la tumeur peut être induite par infection virale.

En fait, les gènes *ras* oscillent à la limite de l'oncogénèse. Presque toute mutation aux positions 12 ou 61 peut convertir un proto-oncogène *c-ras* en un oncogène actif. Aux trois gènes *c-ras* correspond une glycine à la position 12. Si elle est remplacée *in vitro* par l'un des 19 autres acides aminés sauf la proline, le gène *c-ras* muté peut transformer des cellules en culture. A chaque substitution correspond un degré de transformation différent.

La position 61 est occupée par une glutamine dans les gènes *c-ras* de type sauvage. Son remplacement par un autre acide aminé rend généralement le gène potentiellement transformant. Certaines substitutions sont moins efficaces que d'autres ; la proline et l'acide glutamique sont les seuls substituants à n'avoir aucun effet.

Lorsque l'expression d'un gène *c-ras* normal est augmentée, soit parce qu'il est placé sous le contrôle d'un promoteur plus actif, soit parce qu'il est introduit en multiples copies dans les cellules transfectées, les cellules receveuses peuvent être transformées. Certaines lignées tumorales ont également des gènes *ras* amplifiés (voir Tableau 35.6). Une augmentation de 20 fois du taux de protéine *ras*, qui n'est pas habituellement transformante, est suffisante pour permettre la transformation de certaines cellules. L'effet n'a pas été totalement quantifié, mais il amène à la conclusion générale suivante : l'oncogénèse dépend d'une suractivité de la protéine *ras* et peut être due à une augmentation de la quantité de protéine ou (probablement de façon plus efficace) à une variété de mutations qui augmentent l'activité de la protéine.

Ces expériences de transfection ne peuvent être réalisées que sur certains types cellulaires. Bien que les oncogènes transformants aient été isolés à partir de cellules humaines et de rongeurs, la plupart des cibles pour la transformation par des oncogènes sont des cellules de rongeurs en culture. En fait, la différence entre l'origine de l'oncogène (humain) et celle de la cellule receveuse (rongeur) est un facteur important pour permettre au gène donneur d'être distingué sans équivoque de l'ADN receveur. En dépit de cette limitation, la corrélation entre mutation dans les gènes *c-ras* et apparition de tumeurs est suffisamment puissante pour suggérer une relation entre le gène *c-ras* et la capacité d'une tumeur à pousser et disséminer.

MYC AINSI QUE D'AUTRES ONCOGÈNES PEUVENT ÊTRE ACTIVÉS PAR INSERTION, TRANSLOCATION ET AMPLIFICATION

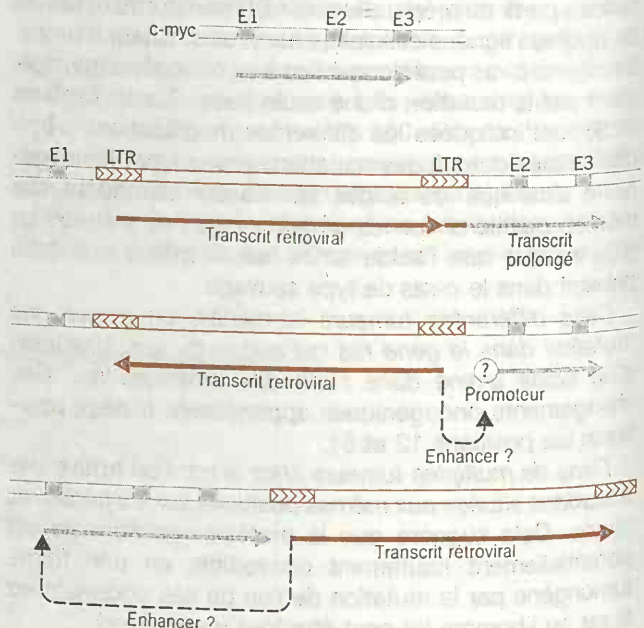
Les proto-oncogènes peuvent être activés par divers événements, qui vont de la mutation de leur séquence codante à la réorganisation des séquences chromosomiques bordantes. Certains proto-oncogènes sont activés par des événements qui modifient leur expression, mais qui laissent la séquence codante intacte. Le mieux caractérisé de ces proto-oncogènes est *c-myc* dont l'expression peut être augmentée par différents mécanismes.

La capacité d'un rétrovirus à transformer une cellule sans exprimer sa séquence *v-onc* fut découverte lors de l'analyse de lymphomes de la bourse dus à la transformation de lymphocytes B par le virus de la leucose aviaire (ALV). Cette même capacité de transformation se rencontre chez le virus de la leucémie murine lors de l'induction de lymphomes de cellules T. Dans chacun des cas, le potentiel transformant du rétrovirus semble être lié à ses LTR plutôt qu'à ses séquences codantes.

Dans de nombreuses tumeurs, le virus a été intégré

Figure 35.5

L'insertion de ALV au locus *c-myc* active le gène et engendre des cellules tumorales.



dans le génome cellulaire à l'intérieur ou près du gène *c-myc*. Ce gène est constitué de trois exons ; le premier est un long exon leader non traduit, et les deux suivants codent pour la protéine *c-myc*. Sur la **Figure 35.5**, sont résumés les différents types d'insertion observés à ce locus.

Les insertions les plus simples à expliquer sont celles qui se sont produites dans le premier intron. Le LTR sert de promoteur et les deux exons codants du gène sont exprimés comme partie intégrante d'un ARN initié par le LTR viral. Placée sous ce nouveau contrôle, la transcription de *c-myc* diffère en deux points de son expression normale : le niveau d'expression est augmenté (car le LTR est un promoteur puissant) et le transcrit a perdu sa séquence leader non traduite.

L'activation de *c-myc* par les deux autres catégories d'insertions ne peut pas être expliquée par la lecture du gène à partir du promoteur viral. Le génome rétroviral peut être inséré à l'intérieur ou en amont du premier intron, mais dans l'orientation inverse du gène, ce qui dirige le promoteur dans la mauvaise direction. Le LTR joue alors probablement le rôle d'un « enhancer » agissant sur un promoteur situé en une position et dans une orientation adéquates en amont des exons codants de *c-myc*. Le génome rétroviral peut également être inséré en aval du gène *c-myc* auquel cas la transcription peut être initiée à partir du promoteur habituel de *c-myc*, mais peut être augmentée par le « enhancer » du LTR rétroviral.

D'autres oncogènes sont activés par l'insertion d'un génome rétroviral, ce sont *c-erbB*, *c-myb*, *c-mos*, *c-H-ras* et *raf*. Plus de dix autres gènes cellulaires (qui n'ont pas été identifiés en tant qu'oncogènes grâce à leur présence dans des virus transformants) près desquels a été mise en évidence une insertion rétrovirale dans certains types de tumeurs, sont présumés oncogènes.

La translocation dans un nouvel environnement est un autre mécanisme par lequel les oncogènes peuvent être activés. Certaines translocations chromosomiques sont associées à l'activation d'oncogènes se situant près des points de cassure impliqués dans l'événement de recombinaison. Ceci fut initialement découvert grâce à la relation qui fut faite entre les translocations affectant les loci codant pour les immunoglobulines et l'apparition de certaines tumeurs. Des translocations chromosomiques spécifiques sont souvent associées à des plasmocytomes chez la souris et à des lymphomes de Burkitt chez l'homme. Ces tumeurs proviennent de lymphocytes B anormaux. Dans les deux espèces, un oncogène porté par un chromosome est mis à proximité d'un locus Ig porté par un autre chromosome (les chromosomes por-

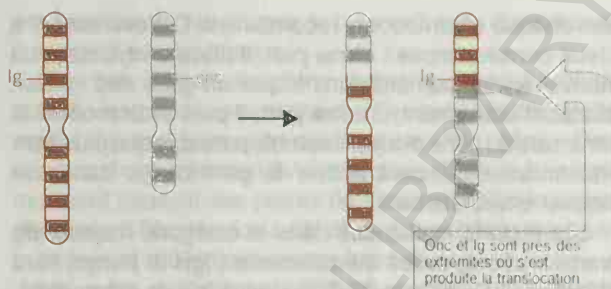


Figure 35.6

Une translocation chromosomique est un événement réciproque qui permet l'échange d'une partie de deux chromosomes. Lorsqu'un gène *onc* est transloqué dans un locus Ig, il peut être activé.

tant les loci Ig sont résumés sur le Tableau 32.1). La nature de la translocation est illustrée sur la **Figure 35.6**.

Chez la souris, la plupart des translocations chromosomiques relient une partie du chromosome 15 à la région du chromosome 12 portant le locus IgH (chaîne lourde des Ig) ; certaines relient la même région du chromosome 15 à la partie du chromosome 6 portant le locus kappa.

Chez l'homme, les translocations concernent le chromosome 8 et le chromosome 14 qui porte le locus IgH ; environ 10 % impliquent le chromosome 8 avec soit le chromosome 2 (au locus kappa), soit le chromosome 22 (au locus lambda).

C'est le même oncogène qui est concerné chez la souris et l'homme, c'est-à-dire *c-myc* qui est situé sur le chromosome 15 murin et le chromosome 8 humain. Lorsque *c-myc* est transloqué au locus des Ig, il est activé. Son activation est l'un des événements qui est associé à la conversion d'une cellule normale à l'état tumoral.

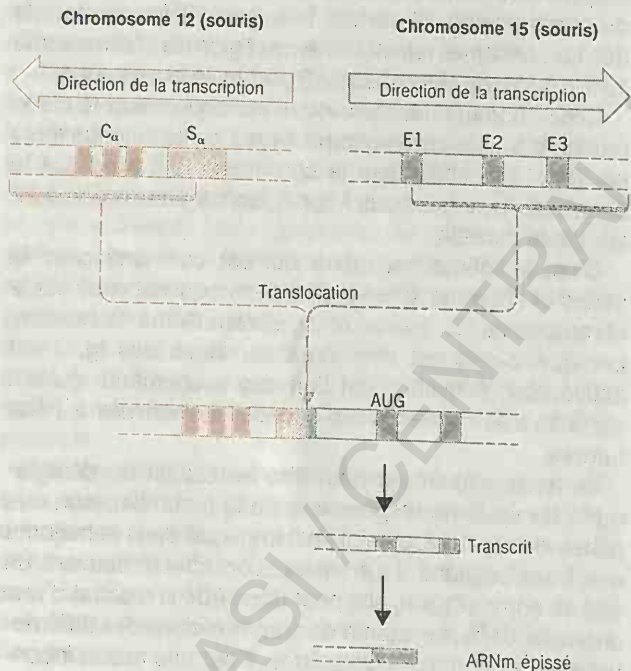
La translocation est peut-être le résultat du dérèglement du système responsable de la recombinaison des gènes des Ig. Le locus IgH est impliqué dans la majorité des translocations. La translocation affecte souvent un site de commutation, elle peut donc être le résultat d'une anomalie dans le système de commutations des différentes classes d'immunoglobulines. Pour une raison inconnue, les enzymes de recombinaison peuvent agir sur un autre chromosome, sur des sites se trouvant au voisinage du gène *c-myc* ainsi que sur des sites localisés près des loci des Ig.

En quoi la translocation active-t-elle le gène *c-myc* ? La translocation ne touche pas des sites fixes mais se produit à différentes positions à l'intérieur d'une région

sur chaque chromosome recombinant. Cet événement a deux conséquences : d'une part, il place *c-myc* dans un nouvel environnement dans lequel un gène des Ig était transcrit activement, d'autre part, il peut avoir modifié la structure du gène *c-myc*. Il semble probable que plusieurs mécanismes puissent activer le gène *c-myc* dans son nouvel environnement.

Les régions impliquées dans la catégorie majeure de translocations, celles qui concernent IgH et *c-myc*, sont décrites sur la **Figure 35.7**. Au locus IgH, la recombinaison se produit le plus souvent à l'intérieur de la région de commutation du gène C_α chez la souris ou du gène C_μ chez l'homme. L'événement de recombinaison se produit dans la région proche du premier exon (non codant) du gène *c-myc*.

Figure 35.7
La plupart des translocations activant *c-myc* se produisent entre la région S du locus des Ig et le premier exon de *c-myc*. Les exons sont représentés par les blocs solides, les introns par les blocs vides.



La région codante de *c-myc* commence dans le second exon, de sorte que la translocation ne modifie pas la séquence codante de la protéine *c-myc*. Cependant, comme le gène perd au minimum la région la plus en amont du premier exon, le gène *c-myc* transloqué doit s'exprimer à partir d'un nouveau promoteur, sa régulation peut donc avoir changé.

La translocation unit le gène *c-myc* au locus des Ig de telle sorte que leur direction de transcription soit divergente. Cela signifie que l'activation du gène *c-myc* n'est pas seulement le résultat de sa transcription à partir d'un des promoteurs des immunoglobulines. Qu'est-ce qui est alors responsable de l'expression du gène *c-myc* transloqué ?

Nous supposons que le gène *c-myc* transloqué n'est pas intact, des séquences se trouvant par hasard en amont du gène peuvent servir de promoteur (la même supposition est faite lors de l'insertion en orientation inverse de rétrovirus dans l'intron). Comme la translocation supprime la région 5' du locus des Ig, le « enhancer » des Ig est perdu dans la majorité des translocations, de telle sorte que nous ne pouvons pas considérer le mécanisme de l'activation du gène transloqué comme étant le résultat de l'augmentation de son expression.

Puisque l'ARNm de *c-myc* tronqué code pour la même protéine que l'ARNm de type sauvage, l'effet transformant ne peut être dû qu'à une ou plusieurs modifications de l'expression du gène. Le niveau d'expression du gène est, dans tous les cas de translocation, augmenté. Cette augmentation varie considérablement selon les tumeurs, elle est de 2 à 10 fois le niveau d'expression normal.

Lorsque seule une partie du gène a été transloquée, l'ARNm *c-myc* est plus court que l'ARNm normal. Il est constitué d'un segment dérivé du premier intron, et des deuxième et troisième exons qui sont épissés ensemble. Afin d'expliquer l'effet transformant, l'on peut supposer que l'absence de la séquence leader non traduite codée par le premier exon augmente la quantité de protéine, soit directement lors de la traduction, soit indirectement par un effet sur l'ARNm lui-même. Une autre hypothèse serait de supposer que le gène de structure n'est plus sous le contrôle d'un régulateur négatif situé habituellement en amont de sa position normale *.

* N. d. T. : Un bloc à l'allongement des transcrits, au-delà de la limite entre la fin du premier exon et du premier intron, a été mis en évidence par de nombreuses équipes au niveau du gène *c-myc*. L'importance de ce bloc est d'être soumis à un contrôle probablement par l'intermédiaire d'une protéine reconnaissant cette région. Une délétion du premier exon de *c-myc* entraîne évidemment la disparition de cette zone ; dans de nombreux cas de réarrangement du gène *c-myc* ne comportant pas de délétion du premier exon, on a démontré que des mutations s'accumulaient avec une grande fréquence dans le premier exon non codant, particulièrement dans sa partie 3'. On peut s'expliquer ce phénomène par une participation du gène *c-myc* transloqué au voisinage d'un locus d'Ig au processus de stimulation des mutations somatiques qui jouent physiologiquement un rôle dans la diversification des Ig. Quoi qu'il en soit, ces mutations pourraient également entraîner une levée du bloc à l'allongement des transcrits, et donc une augmentation de l'expression du gène *myc*. Par ailleurs, la stabilité du messager est également soumise à un contrôle probablement important ; il a été montré dans plusieurs cas que les messagers des gènes *myc* transloqués étaient plus stables que les messagers normaux.

La corrélation qui existe entre l'activation de *c-myc* par insertion ou translocation et le phénotype tumoral suggère que la présence d'une trop grande quantité de protéine *c-myc* est probablement oncogénique. Le potentiel oncogénique de *c-myc* a été démontré directement par la création de souris transgéniques portant le gène *c-myc* normal lié à un enhancer. Ces expériences prouvent que le potentiel oncogénique de *c-myc* n'est pas limité aux cellules lymphoïdes dans lesquelles il est généralement activé.

Les souris transgéniques portant un gène *c-myc* lié à un « enhancer » spécifique des lymphocytes B (enhancer IgH) développent des lymphomes. Les tumeurs sont constituées à la fois de lymphocytes B immatures et matures, ce qui signifie que la surexpression de *c-myc* est tumorigène pour toute la lignée lymphoïde. Les souris transgéniques portant un gène *c-myc* mis sous le contrôle d'un LTR issu du virus murin de tumeur mammaire, développent divers cancers, dont des carcinomes mammaires. Cela suggère que l'expression accrue de *c-myc*

est capable de transformer le type cellulaire dans lequel il apparaît en une tumeur.

Bien que les translocations *c-myc*-Ig soient à l'heure actuelle les mieux caractérisées, elles sont loin de représenter le seul exemple d'activation d'un gène. Des translocations mises en évidence dans d'autres tumeurs montrent souvent des points de cassure associés à des gènes dont l'expression a été altérée, ces gènes sont par conséquent d'éventuels oncogènes.

Un autre cas clair est illustré par le chromosome *Philadelphie* (*PH¹*) présent dans la leucémie myéloïde chronique (LMC). Cette translocation réciproque est trop petite pour être visible sur le caryotype, mais elle lie une région de 5 000 kb située à la fin du chromosome 9 et portant *c-abl* à la région *bcr* du chromosome 22. *bcr* (pour « Breakpoint Cluster Region », région où sont groupés les points de cassure) est définie comme une région d'environ 5,8 kb sur le chromosome 22 dans laquelle se produisent de nombreux points de cassure.

Les conséquences de cette translocation sont résu-

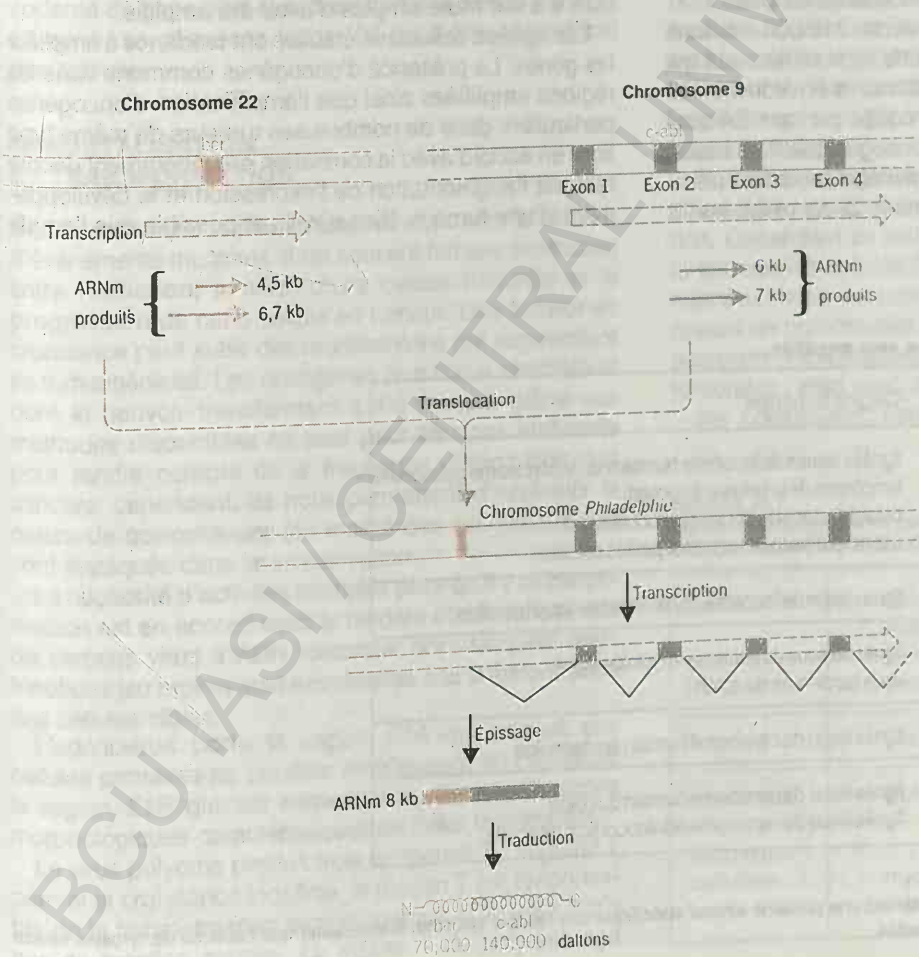


Figure 35.8
La leucémie myéloïde chronique est associée à des translocations entre le chromosome 22 et le chromosome 9 qui engendrent le chromosome *Philadelphie* synthétisant un transcrit de fusion *bcr-c-abl*.

mées sur la **Figure 35.8**. Bien que les points de cassure sur les chromosomes 9 et 22 varient selon les cas, les translocations ont une caractéristique commune : la production d'un transcrit codant pour une protéine de fusion *bcr-abl*, dans laquelle les séquences N-terminales issues de *bcr* sont liées aux séquences *c-abl*.

Le point de cassure sur le chromosome 9 est généralement situé à plus de 15 kb en amont du locus de *c-abl*, ce dernier reste donc intact après la translocation. Par conséquent, une unité de transcription part de *bcr*, s'étend à travers le point de cassure, puis à travers le gène *c-abl* transloqué. La diversité des points de cassure des différentes translocations suggère que la taille du transcrit primaire doit varier significativement. L'ARN qui en résulte relie la partie 5' du transcrit *bcr* au gène *c-abl*. L'épissage du transcrit de fusion donne lieu à un ARNm qui est commun à de nombreux cas de LMC. L'épissage relie un exon de la région *bcr* au second exon du gène *c-abl* : le premier exon de *c-abl* est excisé. La protéine qui en résulte est constituée des séquences *bcr* liées à la majeure partie de la séquence codante de *c-abl*.

La protéine de fusion d'environ 210 000 daltons contient environ 140 000 daltons de la protéine *c-abl* qui est normalement de 145 000 daltons, et environ 70 000 daltons de la protéine inconnue codée par *bcr*. En quoi cette protéine de fusion est-elle oncogénique ? La fusion ou la perte des séquences N-terminales modifient peut-être la conformation des séquences *c-abl* restantes et

activent un potentiel oncogénique latent. Les séquences N-terminales obligent peut-être les séquences *c-abl* à résider en un nouvel endroit de la cellule*.

L'amplification génique est un autre mécanisme par lequel l'expression d'un oncogène peut être augmentée. De nombreuses lignées de cellules tumorales possèdent des régions visibles d'amplification chromosomique comme le montrent les régions de coloration homogène (voir Figure 31.4) ou les chromosomes double-minute (voir Figure 31.5). Dans certains cas, la région amplifiée contient un oncogène connu ou un gène apparenté à un oncogène. Dans d'autres cas, l'utilisation d'une batterie de sondes correspondant à des oncogènes montre que c'est un oncogène particulier qui est amplifié, bien que l'amplification ne soit pas nécessairement visible. Quelques exemples sont résumés sur le **Tableau 35.6**.

Notons que dans la majorité des cas, il est question de lignées cellulaires et non de cultures primaires. Le niveau d'amplification est très variable. Dans aucun cas, nous ne savons si le gène amplifié est de type sauvage ou s'il a été muté en plus d'avoir été amplifié.

Les lignées cellulaires établies ont tendance à amplifier les gènes. La présence d'oncogènes communs dans les régions amplifiées ainsi que l'amplification d'oncogènes particuliers dans de nombreuses tumeurs de même type sont en accord avec la corrélation précédemment décrite qui unit l'augmentation de l'expression et le développement d'une tumeur. Bien sûr, il est possible que l'ampli-

Tableau 35.6
Dans certaines tumeurs, les oncogènes sont amplifiés.

Oncogène	Amplification	Origine tumorale
<i>c-myc</i> " <i>N-myc</i> <i>L-myc</i>	~ 20x 5-10x 5-1000x 10-20x	lignées issues de leucémie humaine et de carcinome pulmonaire lymphome de la bourse du poulet neuroblastome humain et rétinoblastome cancer pulmonaire humain à petites cellules
<i>c-abl</i>	~ 5x	lignée issue de leucémie myéloïde chronique humaine
<i>c-myb</i>	5-10x	lignées issues de leucémie myéloïde aiguë et de carcinome du côlon
<i>c-erbB</i>	~ 30x	lignée issue de carcinome humain épidermoïde
<i>c-K-ras</i> "	4-20x 30-60x	lignée issue de carcinome humain du côlon lignée issue de carcinome adrénocortical de souris

* N. d. T. : La protéine codée par *c-abl* est une protéine kinase spécifique des résidus tyrosine. Il semblerait que l'activité de tyrosine kinase de la protéine hybride *bcr-abl* soit augmentée.

fication du gène favorise le développement d'une tumeur déjà établie, plutôt qu'elle constitue un des événements impliqués dans son initiation.

Deux nouveaux oncogènes apparentés à *c-myc* ont été identifiés grâce à leur présence dans l'ADN amplifié. Les séquences codantes des gènes cellulaires *N-myc* et *L-myc* sont apparentées à une partie du second exon de *c-myc*. L'homologie se limite à des régions relativement courtes. Les gènes codent peut-être pour des protéines ayant une seule fonction en commun — comme, par exemple, la capacité à lier un ligand particulier — et ne forment pas, nécessairement une famille de gènes aux mêmes caractéristiques que la famille des gènes *c-ras* dont les membres sont très ressemblants.

Les oncogènes peuvent être activés par de nombreux mécanismes ; l'oncogène *c-myc* en illustre trois : l'insertion rétrovirale, la translocation chromosomique et l'amplification génique. Le point commun entre ces mécanismes est une augmentation de l'expression de l'oncogène plutôt qu'un changement qualitatif de la séquence codante du gène bien que, dans certains cas, le transcrit ait perdu sa séquence leader (et peut-être régulatrice) non traduite.

IMMORTALISATION ET TRANSFORMATION

Nous savons que la plupart des tumeurs sont le résultat d'événements multiples. Il est souvent fait une distinction entre l'**initiation**, *in vivo*, d'une cellule tumorale et la **progression** de cette cellule en tumeur. Une tumeur en croissance peut subir des modifications qui augmentent sa tumorigénicité. Les oncogènes que l'on a identifiés et dont le pouvoir transformant a été mesuré grâce aux méthodes disponibles ne sont peut-être pas suffisants pour rendre compte de la fréquence d'apparition des cancers, cependant, ils nous permettent d'expliquer la nature de quelques-uns (au moins) des événements qui sont impliqués dans la cancérisation.

La nécessité d'activités multiples pour qu'il y ait cancérisation est en accord avec le modèle établi dans le cas de certains virus à ADN tumoraux pour lesquels deux fonctions (au moins) sont nécessaires à la transformation des cellules cibles.

L'adénovirus porte la région *E1A* qui permet aux cellules primaires de pousser indéfiniment en culture, et la région *E1B* qui est responsable des modifications morphologiques caractéristiques de l'état transformé.

Le virus polyome produit trois antigènes T : le grand T permet la croissance indéfinie, le moyen T est responsable de la transformation morphologique et le petit T n'a pas de fonction connue. Le moyen T de polyome agit

peut-être en interaction avec *c-src* auquel il se lie lors de l'infection.

Les différentes fonctions ne sont pas forcément indépendantes ; le grand T de SV40 allie les fonctions des antigènes grand T et moyen T de polyome et peut à lui seul transformer des cellules.

L'assignation des oncoprotéines de ces différents virus à la fonction d'immortalisation ou de transformation (ou les deux) est résumée dans le **Tableau 35.7**. Selon la classification des fonctions oncogéniques, l'adénovirus E1A pourrait, avec l'aide du moyen T de polyome, transformer des cellules primaires. Cela signifie que les deux fonctions sont nécessaires à la transformation.

Les fonctions de la plupart des oncoprotéines codées par des rétrovirus transducteurs (ou les gènes *c-onc* apparentés) ne sont pas encore caractérisées. Cependant, au vu du comportement des mutants thermo-sensibles, nous savons que *v-src* est nécessaire et suffisant pour établir et maintenir une cellule transformée. *V-src* et le grand T de SV40 sont les seules oncoprotéines connues pour lesquelles une telle faculté ait été définie sans équivoque, bien que dans de nombreux cas la distinction entre protéine immortalisante et protéine transformante soit trouble. Par exemple, E1A possède certaines des fonctions qui sont généralement attribuées aux protéines transformantes.

Plusieurs oncogènes ont été identifiés grâce à leur efficacité à transformer des cellules 3T3 après transfection. Cependant ce test de transformation est limité par le fait que les cellules 3T3 ne sont pas des fibroblastes normaux, mais se sont adaptées à pousser en culture depuis de nombreuses années. En réalité, elles ont subi quelques modifications caractéristiques des cellules tumorales ; elles sont, en particulier, immortelles.

Par conséquent, pour que soit présente une activité

Tableau 35.7
Certains oncoprotéines possèdent les activités immortalisante ou transformante (ou les deux).

Système	Fonction immortalisante	Fonction transformante
SV40	grand T	grand T
Polyome	grand T	moyen T
Adénovirus	E1A	E1B
Rétroviral /cellulaire	myc v-src	ras v-src

oncogénique après transfection, il faut que l'oncogène transfecté soit capable d'induire des modifications morphologiques et autres changements phénotypiques dans une lignée cellulaire établie. Il n'est donc pas surprenant que les résultats obtenus après les tests de transfection soient biaisés ; en effet, ils montrent que seulement 10 à 20 % des tumeurs humaines spontanées possèdent de l'ADN ayant une activité transformante. Cela signifie peut-être que le test ne permet la détection que d'un type particulier d'oncogène, capable d'effectuer la transformation de cellules immortelles. Les principaux produits décelés par les tests de transfection dans les cellules 3T3 sont les gènes *c-ras* mutés. Ils n'ont pas la propriété de transformer des cellules primaires (cellules prélevées directement à partir de l'animal) *in vitro*, ce qui est en accord avec l'hypothèse suggérant qu'ils ne peuvent transformer que des cellules déjà immortalisées.

Comment doit-on préparer une cellule pour qu'il soit possible de la transformer par un oncogène *ras* ? Les protocoles utilisés afin d'établir des fibroblastes en culture nécessitent essentiellement l'adaptation des cellules à la croissance indéfinie. Nous ne comprenons pas les changements par lesquels elle s'opère, mais nous pouvons nous demander s'ils font intervenir les fonctions impliquées dans l'immortalisation.

Quelles que soient les fonctions nécessaires à l'immortalisation, elles peuvent être apportées (ou circonvenues) par d'autres oncogènes. Bien que la transformation de fibroblastes primaires ne soit pas réalisable par les oncogènes *ras* seuls, elle peut être obtenue par double transfection de *ras* et d'un autre oncogène. Plusieurs oncoprotéines sont responsables de la fonction immortalisante, ce sont *c-myc*, E1A de l'adénovirus et le grand T de polyome. Ce groupe d'oncogènes confère un phénotype identique à celui résultant des protocoles d'immortalisation. Un autre groupe, dont font partie *ras*, E1B de l'adénovirus et le moyen T de polyome, induit les modifications qui caractérisent le phénotype transformé.

Aucun trait commun particulier n'est évident entre *ras* et E1A et entre *myc* et E1B. Nous ne savons pas si ces oncoprotéines ont la même fonction, activent la même voie de synthèse, ont le même résultat final. Nous ne savons pas non plus comment les autres oncoprotéines entrent dans ce schéma, si elles peuvent être classées dans l'un de ces groupes ou si elles ont des actions complètement différentes.

Dans aucun cas, il n'a pu être mis en évidence une relation entre les activités moléculaires connues de certai-

nes oncoprotéines et leurs capacités à transformer ou immortaliser des cellules *. Les oncoprotéines des virus à ADN tumoraux SV40 et adénovirus ont des effets régulateurs sur la réplication ou la transcription ; il est bien sûr tentant de supposer que ce sont eux qui sont responsables de l'oncogénicité. Cependant, dans chacun des cas, ces effets peuvent être séparés de la capacité à affecter le phénotype cellulaire. L'antigène grand T de SV40 est essentiel pour le cycle de vie viral, sa liaison à l'origine de l'ADN viral est nécessaire car elle permet à la réplication d'être initiée. Cependant, des mutations qui éliminent l'activité de liaison à l'ADN n'empêchent pas l'antigène T d'immortaliser et de transformer les cellules. Le(s) produit(s) E1A de l'adénovirus active(nt) certains promoteurs viraux et cellulaires. Une fois encore, lorsque cette activité est mutée, la capacité d'immortalisation n'est pas perdue. Par conséquent, ces oncoprotéines possèdent probablement différents domaines responsables de leurs actions moléculaires et de leurs effets immortalisants/transformants, ces derniers demeurant non caractérisés en termes d'activités protéiques.

LES FONCTIONS POSSIBLES DES ONCOPROTÉINES

A l'heure actuelle, nous ne pouvons définir aucun mécanisme commun à tous les gènes *onc* grâce auquel leur potentiel oncogénique est activé. Une grande variété de mécanismes activant les proto-oncogènes est résumée dans le **Tableau 35.8**.

Certains proto-oncogènes nécessitent, pour devenir oncogéniques, des changements spécifiques de séquence, comme le montre la différence qui existe entre *v-src* (oncogénique à taux élevé) et *c-src* (non oncogénique à taux élevé). D'autres peuvent être faiblement oncogéniques lorsque le *c-onc* est exprimé à des taux augmentés, mais les changements de séquences trouvés dans *v-onc* ou apparaissant spontanément dans *c-onc* peuvent activer le gène de façon plus effective, comme c'est le cas pour *ras* et peut-être pour *mos*.

Certains oncogènes peuvent être activés par un changement environnemental. La séquence de *v-abl* est oncogénique dans les cellules lymphoïdes lorsqu'elle est attachée à la séquence virale *gag*, mais ne l'est pas lorsqu'elle en est séparée. Peut-être cela a-t-il un rapport avec le fait que *c-abl* est activé lorsqu'il est fusionné à *bcr* par translocation (voir Figure 35.8).

* N. d. T. : En réalité, certains oncogènes codent pour des équivalents plus ou moins modifiés de facteurs de croissance ou de leur récepteur (voir plus loin). Dans ce cas, le mécanisme de la transformation cellulaire est très évidemment relié à une perpétuation anormale de la stimulation de la prolifération cellulaire. Les oncogènes *v-sis* (équivalent du facteur de croissance PDGF), *v-fms* (équivalent du récepteur du facteur de croissance CSF-1) et *v-erbB* (équivalent du récepteur du facteur de croissance EGF) en sont des exemples.

Tableau 35.8
L'oncogénicité peut être due à une modification quantitative ou qualitative des gènes *c-onc*.

Onco-protéine	Oncogénicité de <i>c-onc</i> en excès	Modifications de la séquence oncogénique
src	aucune	modification de la séquence dans <i>v-src</i>
abl	faible ?	liaison à la séquence gag dans <i>v-abl</i> ou substitution N-terminale dans <i>c-abl</i>
erbB	?	?
neu	?	mutation ponctuelle dans le domaine transmembranaire
mos	faible	séquence C-terminale de <i>c-mos</i>
ras	faible	modifications des codons 12, 59, 61
myc	importante	pas nécessaire

D'autres gènes *c-onc* sont activés de façon effective uniquement grâce à l'augmentation de leur expression (ou peut-être à une altération de leur régulation) ; *c-myc* en est le paradigme.

Que les oncogènes soient activés par des modifications qualitatives ou quantitatives, l'on peut supposer qu'ils peuvent influencer (directement ou indirectement) les fonctions liées à la croissance cellulaire. Les cellules transformées n'ont pas les restrictions de croissance imposées aux cellules normales, comme par exemple la dépendance au sérum ou l'inhibition de contact inter-cellulaire. Nous supposons que les oncogènes, individuellement ou de concert, sont responsables d'une série de

modifications phénotypiques associées avec l'état transformé.

Quelles fonctions les produits des proto-oncogènes remplissent-ils dans la cellule normale, et en quoi celles-ci sont-elles différentes dans la cellule transformée ? Dans la plupart des cas, nous n'avons aucune réponse. Certaines caractéristiques biochimiques sont communes à plusieurs oncogènes, mais dans aucun cas, nous ne connaissons la relation qui existe entre l'activité biochimique et la capacité de transformation (voir N. d. T. précédente). Des informations concernant la localisation des produits protéiques ainsi que leurs activités biochimiques sont résumées dans le **Tableau 35.9**.

Tableau 35.9
Les oncoprotéines ont des fonctions et des localisations diverses.

Gène	Protéine c-onc		Protéine v-onc		Fonction
	Localisation	Taille	Taille	Composants	
<i>src</i>	membranes	60K	60K	src	tyrosine kinase
<i>abl</i>	cytoplasme	150K	120	gag-abl	tyrosine kinase
<i>erbB</i>	membrane plasmique	133K	142K	P68-erbB	tyrosine kinase
<i>neu</i>	membrane plasmique	185K	—	—	tyrosine kinase
<i>fps</i>	soluble	98K	140K	gag-fps	tyrosine kinase
<i>fms</i>	membrane plasmique	170K	140K	fms	tyrosine kinase
<i>mos</i>		37K	37K	env-mos	thréonine kinase
<i>ras</i>	membranes	21K	21K	ras	" & GTPase
<i>myc</i>	noyau	58K	} 110K	} myc	liaison à l'ADN
				} gag-myc	
<i>myb</i>		>75K	48K	myb	liaison à l'ADN

Une différence notable entre les protéines codées par *c-onc* et celles codées par *v-onc* est que la séquence portée par le virus peut être exprimée par lecture à partir de la séquence virale qui la précède. Dans ce cas, la séquence de la protéine *v-onc* fait partie d'une protéine plus grande qui porte également des fonctions virales (souvent, la séquence gag qui se lie à l'ARN viral dans le cœur de la particule virale). Src est une exception, car la protéine virale est synthétisée indépendamment, et ressemble, par conséquent, à la protéine cellulaire.

Plusieurs protéines sont localisées sur les membranes, membrane cytoplasmique et membrane nucléaire, d'autres sont solubles dans le cytoplasme, et d'autres encore sont localisées dans le noyau. Aucune caractéristique commune entre ces protéines n'est évidente.

Il est possible que des modifications dans la localisation de la protéine *onc* puissent être associées avec le potentiel transformant. Le virus REV-T peut infecter les cellules lymphoïdes ainsi que les fibroblastes, mais ne transforme que les premières. Son oncoprotéine, *v-rel*, est localisée dans le cytoplasme des cellules lymphoïdes transformées, alors qu'elle se trouve dans le noyau dans les fibroblastes (non transformés).

Nombreux sont les produits des gènes *onc* qui partagent une fonction catalytique ; c'est l'activité de **tyrosine kinase**. Cette activité enzymatique ajoute des groupes phosphates aux résidus tyrosine dans les protéines. Les protéines ayant cette activité possèdent un domaine reconnaissable, d'environ 30 résidus, qui contient le centre catalytique. Généralement, au moins une partie des molécules protéiques sont localisées sur la membrane plasmique.

La première oncoprotéine de ce type à avoir été caractérisée est *v-src*, qui, en plus d'agir sur d'autres protéines, peut s'autophosphoryler. Afin d'identifier une fonction de phosphorylation qui jouerait un rôle dans la transformation cellulaire, les recherches se sont concentrées sur l'identification de substrats cellulaires pouvant être la cible de *v-src*, sans que ceux-ci soient reconnus par *c-src*. Cependant, aucun substrat n'a encore été mis en relation avec la cause de la transformation.

La protéine *v-src* doit être capable de s'attacher à la membrane afin d'exercer sa fonction oncogénique. La protéine est myristylée à son extrémité aminée terminale ; cette modification lui permet de s'attacher à la membrane.

Une activité de thréonine kinase est également trouvée

dans les protéines *v-onc* (comme par exemple, *mos* ou *raf*) ; la signification de sa présence est une fois encore peu claire. La présence d'une thréonine à la position 59 permet à la protéine *v-ras* d'être phosphorylée (ce n'est pas le cas de *c-ras* ; voir Tableau 35.5). Les protéines *ras* ont une activité GTPasique, mais elle ne semble pas essentielle à leur potentiel oncogénique*.

Virtuellement, tous les proto-oncogènes sont transcrits, au cours du développement normal, à des taux relativement bas. Outre le fait que ces gènes remplissent certainement des fonctions cellulaires, leur mode d'expression (généralement mesuré par leur transcription) n'a pas encore permis de dévoiler des indications quant à la nature de leur fonction.

Dans un certain nombre de cas, une homologie de séquence révèle une identité ou une relation avec un gène connu, comme le résume le **Tableau 35.10**.

La caractéristique frappante qui relie ce groupe de gènes est qu'ils codent tous pour des récepteurs ou des facteurs impliqués dans le développement d'un type particulier de cellule. Comme l'indique son nom, le récepteur à l'EGF est localisé dans la membrane plasmique, où il est responsable de la réponse au facteur de croissance EGF. Le facteur de croissance dérivé des plaquettes (PDGF : « platelet derived growth factor ») consiste en deux chaînes étroitement apparentées (dérivant peut-être de l'expression alternative d'un gène unique) ; il fait partie d'une famille de facteurs de croissance dont chaque membre est spécifique d'un type cellulaire. Le récepteur CSF-I sert de médiateur dans l'action du facteur I de stimulation des colonies, un facteur de croissance macrophagique qui stimule les précurseurs hématopoïétiques afin qu'ils forment des phagocytes mononucléés. Les récepteurs ont une activité

Tableau 35.10

Certains oncogènes viraux sont apparentés à des gènes codant pour des facteurs de croissance cellulaire spécifiques ou leurs récepteurs.

Gène	Produit cellulaire	Fonction du produit cellulaire
<i>c-erbB</i>	récepteur à l'EGF	stimule la croissance cellulaire épidermique
<i>c-sis</i>	chaîne B du PDGF	stimule la croissance des fibroblastes
<i>c-fms</i>	récepteur au CSF-I	développement des cellules phagocytaires mononucléées au cours de l'hématopoïèse

* N. d. T. : Quoique les résultats concernant la relation entre l'activité GTPasique du produit *p21^{ras}* et son pouvoir transformant aient fait l'objet d'études quelque peu contradictoires, on s'accorde néanmoins aujourd'hui pour reconnaître qu'il y a bien une relation entre activité GTPasique et pouvoir transformant. Cela pourrait parfois d'ailleurs être indirect : une protéine GAP (« GTPase activating protein ») contrôle normalement l'hydrolyse en GDP du GTP fixé sur *p21^{ras}*. Des mutations perturbant l'interaction *p21^{ras}*-GAP pourraient stimuler le pouvoir transformant de *ras* en retardant l'hydrolyse du GTP, c'est-à-dire probablement l'inactivation du complexe *p21^{ras}*-GAP.

tyrosine kinase, et leur fonction est impliquée dans la phosphorylation de substrats cellulaires.

La recherche d'homologies de séquences indique que les gènes *c-erbB* et *c-sis* codent respectivement pour les récepteurs à l'EGF et au PDGF-B. Le gène *c-fms* comporte, ou du moins est très étroitement apparenté au gène codant pour le récepteur CSF-I. Un autre oncogène dérivé d'un récepteur est l'oncogène *neu* ; il n'a pas d'équivalent *v-onc* et a été détecté par transfection dans les cellules 3T3 (voir Tableau 35.4). La séquence de *neu* est apparentée à celle de *erbB* — *neu* est parfois appelé *erbB2* — et par conséquent, code probablement pour le récepteur d'un autre facteur de croissance (actuellement non identifié)*.

Dans aucun de ces cas, nous ne connaissons les différences qui existent entre l'oncogène et le proto-oncogène. L'oncogène *v-erbB* est une version tronquée de *c-erbB*, qui a retenu les domaines transmembranaires comportant la tyrosine kinase, mais qui a perdu la moitié N-terminale de la protéine qui pourrait être impliquée dans d'autres fonctions effectrices. D'autre part, l'oncogène *neu*, détecté par transfection dans les cellules 3T3, diffère uniquement du proto-oncogène par une mutation ponctuelle située dans le domaine transmembranaire ; à ce que nous savons, la mutation n'affecte ni la localisation ni l'activité tyrosine kinase. Il semble plausible qu'un mauvais fonctionnement d'une réaction de régulation, qui serait déclenché par un facteur de croissance, puisse être la cause de la transformation, mais nous n'avons encore aucune information sur la nature de ce mauvais fonctionnement**.

Des interactions entre oncogènes ou proto-oncogènes peuvent jouer un rôle important. Par exemple, PDGF stimule l'expression de plusieurs gènes dont font partie les proto-oncogènes *c-myc* et *c-fos*.

Plusieurs proto-oncogènes possèdent des fonctions suffisamment identiques pour avoir été conservées lors de l'évolution. Si l'on s'étend des mammifères aux mouches, l'on constate que les séquences *c-src*, *c-ras* et *c-abl* ont toutes des homologues chez *Drosophila*. Des proto-oncogènes d'insectes ont été caractérisés, mais

nous n'avons pas encore de données sur les propriétés des mutants.

La levure *S. cerevisiae* possède deux gènes *ras*, apparentés à *H-ras*, mais toutefois plus grands. Leur produit stimule l'enzyme adénylate cyclase, qui catalyse la formation d'AMP cyclique, une petite molécule régulatrice bien caractérisée.

Nous ne savons pas comment les fonctions des protéines *ras* de levure sont apparentées à l'oncogénicité de celles des eucaryotes supérieures, mais ces propriétés consolident l'analogie qui existe entre les protéines *ras* et les **G protéines**, qui transmettent des signaux à partir des récepteurs sur la surface des cellules de mammifère. Les G protéines sont constituées de trois sous-unités ; quand elles sont activées, elles se dissocient et la sous-unité alpha active l'adénylate cyclase, puis s'inactive en hydrolysant le GTP. Les protéines *ras* et la G protéine alpha ont par conséquent deux caractéristiques communes : la même localisation et la même fonction biochimique.

Les protéines *ras* de levure et de mammifères ont, au moins à un certain point, une fonctionnalité interchangeable. Les fonctions des protéines *ras* de levure peuvent être substituées par *v-ras* ou *c-ras*. La protéine *ras* mutante de levure est oncogénique lors des tests de transfection. La possibilité de soumettre la levure à des mutations intensives peut ainsi permettre de bien définir que ce sont les fonctions des gènes *ras* de *S. cerevisiae*. Sont-ils impliqués dans la croissance cellulaire ? Les anomalies dans les mutants de levure vont-elles mettre en évidence des analogies avec des cellules de mammifères ?

La conservation de proto-oncogènes dans divers organismes peut ouvrir des horizons quant à l'identification de leurs fonctions cellulaires normales. A l'heure actuelle, cependant, nous ne pouvons établir une relation entre ce que nous savons sur les gènes *c-onc* et les activités de leurs dérivés transformants. Nous soupçonnons « l'immortalisation » d'être le fruit de plus d'un type de modification. Nous possédons des marqueurs phénotypiques de la « transformation », mais nous ne comprenons pas

* N. d. T. : De nouvelles classes d'oncogènes ont été identifiées ces dernières années : *c-erbA* équivalent de *v-erbA* du rétrovirus de l'érythroblastose aviaire code en fait pour le récepteur des hormones thyroïdiennes. L'oncogène HAP, activé par insertion du virus de l'hépatite B au cours du développement d'un hépatocarcinome, code pour un récepteur de l'acide rétinolique. Ces deux récepteurs appartiennent à la superfamille des récepteurs nucléaires qui comprend aussi les récepteurs des hormones stéroïdes et de la vitamine D. Ce sont des protéines se liant à l'ADN et modulant la transcription des gènes adjacents. L'oncogène *v-jun* est l'équivalent du gène cellulaire *c-jun* qui code pour le facteur transcriptionnel Ap-1. Ap-1 appartient à la famille des facteurs de transcription ; il est notamment impliqué dans la réponse transcriptionnelle de certains gènes à la stimulation par des agents mitogènes.

** N. d. T. : Les récepteurs tronqués du type de *v-erbB* pourraient transmettre le signal de prolifération en l'absence même de ligand, c'est-à-dire des facteurs de croissance. Il a été démontré que *c-fms*, récepteur du facteur CSF-I, s'autophosphorylait sur une tyrosine au niveau de l'extrémité C-terminale. Cette phosphorylation serait un mécanisme de rétrocontrôle inactivant la transmission du message mitotique. *V-fms* diffère de *c-fms* au niveau de son extrémité C-terminale qui ne comporte pas d'équivalent de cette tyrosine, cible de l'autophosphorylation.

sa base génétique (ou épigénétique). Nous ne savons pas comment est acquise, *in vivo*, la capacité à métastaser.

Les corrélations entre l'activation d'oncogènes et la croissance de tumeurs sont importantes dans certains cas, mais toujours est-il que la nature de l'événement initiateur reste inconnue. Il semble clair que l'activité oncogénique aide à la croissance de la tumeur, mais cette activation peut apparaître (et être sélectionnée dans ce but) après l'événement d'initiation et au cours du début de la croissance de la tumeur. Nous espérons que la connaissance des fonctions des gènes *c-onc* permettra d'avoir un aperçu de la régulation de la croissance cellulaire dans les cellules normales autant qu'anormales, afin qu'il devienne possible de définir les événements nécessaires à l'initiation et à l'établissement de tumeurs.

POUR EN SAVOIR PLUS

Les oncogènes cellulaires et viraux ont été récapitulés par Bishop (*Ann. Rev. Biochem.* **52**, 301-354, 1983 ; *Cell* **42**, 23-38, 1985). L'activation des oncogènes a été analysée par Varmus (*Ann. Rev. Genet.* **18**, 553-612, 1984). La relation qui existe entre les translocations impliquant les loci des Ig et la formation de tumeurs a été passée en revue par Klein (*Cell* **32**, 311-315, 1983) et Perry (*Cell* **33**, 647-649, 1983). L'induction de tumeurs par des oncogènes chez les souris transgéniques a été observée en premier par Stewart, Pattengale & Leder (*Cell* **38**, 627-637, 1984). Les relations qui existent entre les produits des oncogènes et les facteurs de croissance ont été résumées par Heldin & Westermark (*Cell* **37**, 9-20, 1984). Les tyrosine kinases ont fait l'objet d'une revue par Hunter & Cooper (*Ann. Rev. Biochem.* **54**, 897-930, 1985).

N. d. T. : Les « anti-oncogènes ».

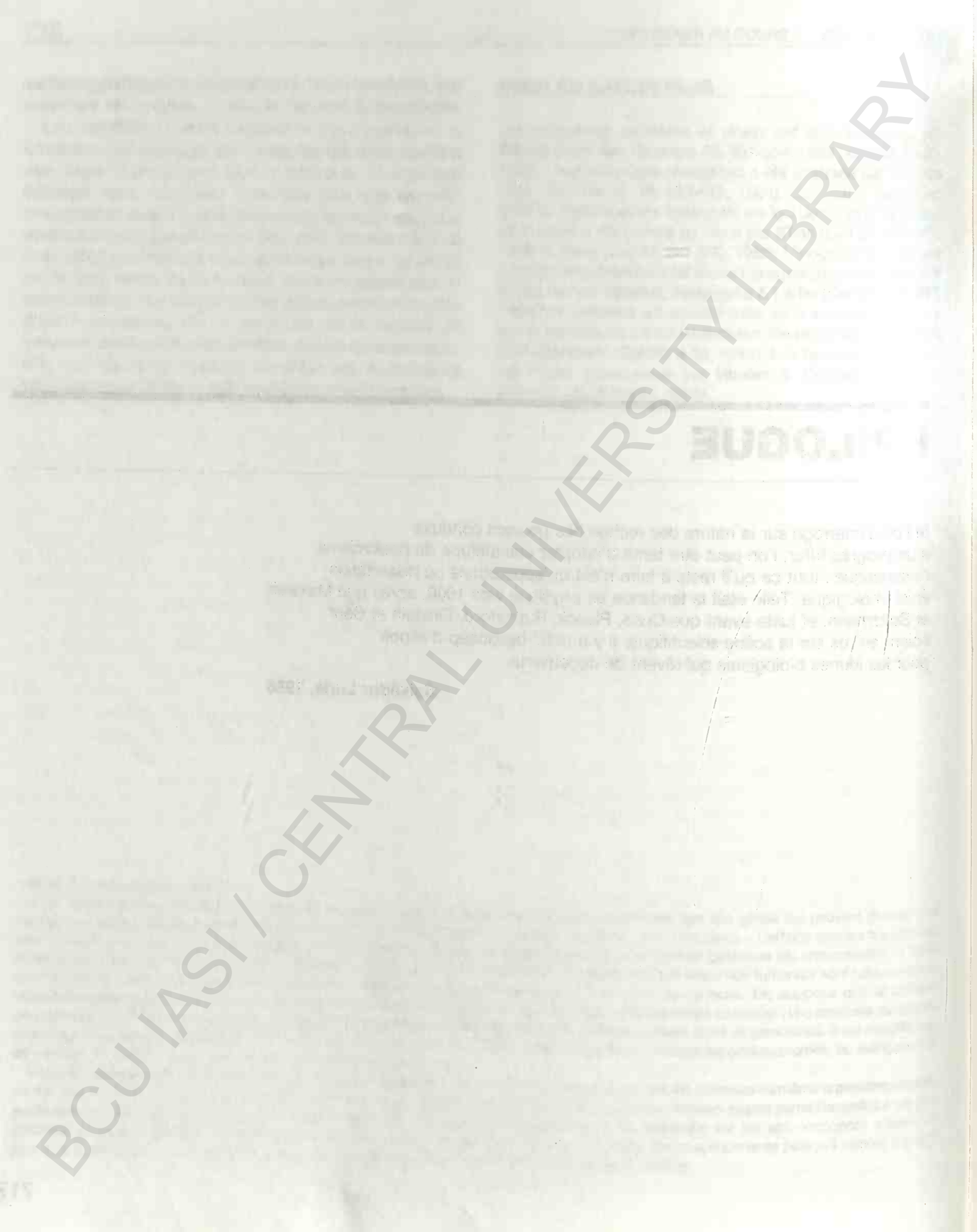
Il est apparu depuis 1986 que la prolifération cellulaire n'était pas seulement contrôlée positivement (par des gènes qui peuvent devenir des oncogènes), mais aussi négativement par des substances anti-prolifératives, produits de véritables « anti-oncogènes ». Certains cancers héréditaires sont associés à des anomalies cytogéniques ou infra-cytogéniques, qui aboutissent à une perte de matériel génétique (du chromosome 11 dans le néphroblastome, ou tumeur de Wilms ; du chromosome 13 dans le rétinoblastome). Habituellement, les tissus non tumoraux sont hétérozygotes pour la délétion d'un locus particulier, les cellules tumorales étant homozygotes pour l'altération de ce locus. On suppose que la délétion constitutionnelle transmise comme un caractère mendélien à la descendance (s'il y a !) explique la prédisposition au cancer ; une anomalie surajoutée de l'autre locus provoquera la tumeur. Le gène de prédisposition au rétinoblastome (Rb) a été récemment cloné et caractérisé. Il est modifié non seulement dans les rétinoblastomes, mais aussi dans d'autres tumeurs (d'origine mésenchymateuse comme les ostéosarcomes, ou autre comme les cancers du sein et les cancers du poumon à petites cellules).

Un phénomène de perte de matériel génétique dans les cellules d'un tissu cancéreux comparé aux cellules normales du même organisme semble, en fait, très fréquent, voire la règle ; cela laisse supposer que le développement d'un cancer comporte plusieurs étapes parmi lesquelles il y a non seulement l'activation d'oncogènes, mais aussi la perte ou l'inactivation d'anti-oncogènes. La recherche sur les anti-oncogènes s'annonce théoriquement aussi fructueuse que celle sur les oncogènes ; sur le plan pratique, elle pourrait même être plus importante puisqu'il semble (a priori) plus aisé d'administrer, pour guérir un cancer, une substance manquante que d'inactiver un gène modifié.

ÉPILOGUE

Si l'on s'interroge sur la nature des recherches pouvant conduire à un progrès futur, l'on peut être tenté d'adopter une attitude de pessimisme romanesque : tout ce qu'il reste à faire n'est qu'applications ou dissertation épistémologique. Telle était la tendance en physique vers 1900, après que Maxwell et Boltzmann, et juste avant que Curie, Planck, Rutherford, Einstein et Bohr soient entrés sur la scène scientifique. Il y a donc beaucoup d'espoir pour les jeunes biologistes qui rêvent de découverte.

Salvador Luria, 1986



BCU IASI / CENTRAL UNIVERSITY LIBRARY

BIBLIOQUE

DÉCOUVERTES MAJEURES QUI ONT CHANGÉ NOTRE VISION DE LA GÉNÉTIQUE

Si l'on regarde en arrière à partir du point de vue avantageux que représentent les années 1980, nous pouvons voir les changements qui se sont opérés depuis que Avery montra en 1944 que l'ADN était le matériel génétique. Quelques-unes des années décisives pour les découvertes d'importance majeure, que ce soit d'un point de vue intellectuel ou pratique, furent :

- 1948 Beadle et Tatum montrent qu'un gène code pour une seule protéine.
- 1953 Watson et Crick proposent la structure en double hélice de l'ADN.
- 1958 Meselson et Stahl démontrent que l'ADN se réplique de façon semi-conservative.
- 1961 Découverte de la nature en triplet du code génétique.
L'ARN messager est découvert.
Jacob et Monod proposent le modèle de l'opéron comme modèle de régulation d'un gène.
- 1970 Temin et Baltimore rapportent la découverte de la transcriptase inverse dans les rétrovirus.
- 1974 Les gènes eucaryotes sont clonés dans des plasmides bactériens.
- 1977 Le séquençage de l'ADN devient possible.
Les gènes morcelés sont découverts et l'épissage de leurs transcrits est inféré.
- 1981 L'activité catalytique de l'ARN est découverte.

Des souris et des mouches transgéniques sont obtenues par introduction d'un nouvel ADN dans la lignée germinale.

Le rythme des découvertes reste actif. Les découvertes qui nous apparaissent majeures aujourd'hui ont certainement des conséquences plus restreintes que celles des découvertes majeures effectuées au début de cette période ; nous nous intéressons maintenant à montrer que l'ADN peut être réarrangé ou transfecté plutôt qu'à élucider sa structure fondamentale ou son mode de réplication. A l'heure actuelle, il n'y a pas encore de fin en vue pour les découvertes, intellectuelles et pratiques, concernant l'ADN et son expression.

Nous avons parcouru un long chemin depuis l'établissement du concept selon lequel le gène était un locus isolé et fixe codant pour une protéine unique. Les gènes peuvent être localisés dans de grands groupes contenant des séquences apparentées. Ils peuvent être exprimés alternativement en plusieurs versions de protéines. Ils peuvent être reconstruits par réarrangement physique de séquences au cours du développement somatique. De nouveaux gènes peuvent même être ajoutés à la lignée germinale par introduction d'ADN dans les œufs. Semblant chaque fois en contradiction avec les conceptions du moment, chaque découverte a, en fait, été rapidement assimilée grâce à un élargissement de notre conception du gène et de ses capacités.

Les questions les plus pressantes du moment concernent l'expression des gènes, et en particulier la nature du réseau des événements régulateurs qui doivent être

impliqués dans le développement d'un embryon en un eucaryote adulte. Les gènes sont-ils contrôlés par une cascade de réactions temporelle ? Quelles sont les distinctions géographiques entre les différentes parties de l'organisme qui sont établies initialement, puis perpétuées et étendues ? Jusqu'où, en réalité, les réseaux impliqués dans la régulation des gènes expliquent-ils le processus de développement ?

Nos idées sur des solutions possibles à ces problèmes sont certainement plus diffuses qu'elles ne l'étaient il y a vingt ou trente ans — il nous manque la base pour supposer qu'il y a un code de développement unique à déchiffrer, analogue à celui du code génétique. Les découvertes fondamentales restent encore à faire et nous devons nous demander quelles seront les prochaines doctrines génétiques à être bouleversées.

GLOSSAIRE

Abondance d'un ARNm : nombre moyen de molécules par cellule.

Activateur (enhancer) : séquence agissant en *cis* qui augmente l'utilisation de (certains) promoteurs eucaryotes et qui peut agir quelle que soit l'orientation ou sa localisation (amont ou aval) par rapport au promoteur.

ADNase (désoxyribonucléase) : enzyme qui attaque les liaisons dans l'ADN.

ADN à séquence simple : ADN satellite.

ADNc : ADN simple brin, complémentaire d'un ARN à partir duquel il est synthétisé par transcription inverse *in vitro*.

ADN cœur : les 146 pb d'ADN contenues dans la particule cœur du nucléosome.

ADNct : ADN du chloroplaste.

ADN de liaison (linker) : ADN contenu dans un nucléosome en plus de l'ADN cœur de 146 pb.

ADN double brin hétérologue (hybride) : ADN engendré par l'appariement de simples brins complémentaires dérivés de différentes molécules double brins parentales ; ils apparaissent durant la recombinaison génétique.

ADN hautement répétitif : c'est le premier composant à se réassocier et il peut être assimilé à l'ADN satellite.

ADN hybride : voir ADN double brin hétérologue.

ADNmt : ADN mitochondrial.

ADN non répétitif : ADN dont les cinétiques de réassociation sont celles attendues pour des séquences uniques.

ADN répétitif : se comporte dans la réaction de réassociation comme si de nombreuses séquences (apparentées ou identiques) étaient présentes dans le composant, permettant à toute paire de séquences complémentaires de se réassocier.

ADN replié : ADN renaturé à partir d'ADN dénaturé par réassociation intrabrin de répétitions inversées.

ADN satellite : ADN constitué de nombreuses répétitions en tandem (identiques ou apparentées) d'une courte unité de base.

ADN se liant au temps zéro : ADN trouvé sous forme double brin au début de la réaction de réassociation ; il résulte de la réassociation intra-moléculaire de séquences répétées inversées.

Allèle : l'une des formes alternatives d'un gène occupant un locus donné sur un chromosome.

Allèle codominant : allèles contribuant également au phénotype. Aucun n'est dominant sur l'autre.

Allèle récessif : allèle masqué dans le phénotype d'un hétérozygote par l'allèle dominant ; il est souvent dû à l'inactivité ou à l'absence du produit de l'allèle récessif.

Allèle récessif létal : allèle qui est létal lorsque la cellule est homozygote pour cet allèle.

α -amanitine : octapeptide bicyclique dérivé du champignon vénéneux *Amanita phalloide*. Inhibe la transcription par certains ARN polymérases eucaryotes, spécialement l'ARN polymérase II.

Aminoacyl ARNt : ARN de transfert transportant un acide aminé ; la liaison covalente s'effectue entre le groupe NH₂ de l'acide aminé et les groupes 3' ou 2'OH de la base terminale de l'ARNt.

Aminoacyl ARNt synthétase : enzyme responsable de la liaison covalente des acides aminés aux positions 2' ou 3'OH de l'ARNt.

Amorce : courte séquence d'ARN ou d'ADN qui est appariée à un acide nucléique simple brin et qui apporte une extrémité 3'-OH libre à partir de laquelle une polymérase (ADN polymérase ou transcriptase inverse) commence la synthèse d'une chaîne déoxyribonucléotidique.

AMP cyclique : molécule d'AMP dans laquelle le groupe phosphate est lié à la fois aux extrémités 3' et 5' d'un ribose ; sa liaison active la CAP, régulateur positif de la transcription des procaryotes.

Amplification : se réfère à la production de copies supplémentaires d'une séquence chromosomique trouvée sous forme d'ADN extra ou intra-chromosomique.

Animaux transgéniques : animaux créés par l'introduction de nouvelles séquences d'ADN dans la lignée germinale.

Anneau de Balbiani : nodule (puff) extrêmement volumineux sur une bande d'un chromosome polytène.

Anticorps : protéine (immunoglobuline) produite par les lymphocytes B qui reconnaît un antigène étranger particulier et déclenche la réponse immunitaire.

Antigène : toute molécule dont l'entrée dans l'organisme provoque la synthèse d'un anticorps (immunoglobuline).

Antigène de transplantation : protéine codée par un locus majeur d'histocompatibilité, présente dans toutes les cellules de mammifères et impliquée dans l'interaction entre les lymphocytes.

Antiparallèle : les brins de la double hélice sont orientés de façon antiparallèle de telle sorte que l'extrémité 5' d'un brin se trouve face à l'extrémité 3' de l'autre brin.

Apo-inducteur : protéine qui se lie à l'ADN pour initier la transcription par l'ARN polymérase.

Appariement : association de simples brins complémentaires d'ADN pour former une double hélice.

Appariement de chromosomes : voir synapse.

Archéobactérie : lignée mineure de procaryotes qui peuvent avoir des introns dans le génome.

ARNase : enzyme de dégradation de l'ARN.

ARNc : n'importe lequel des petits ARNm cytoplasmiques, molécules qui sont présentes dans le cytoplasme et (parfois) dans le noyau.

ARNm abondants : petit nombre d'espèces individuelles d'ARNm, chacun présent à raison de multiples copies par cellule.

ARNm monocistronique : code pour une protéine.

ARNm peu abondant (complexe) : ensemble constitué d'un grand nombre d'espèces d'ARNm, chacune présente en très peu de copies par cellule.

ARNm polycistronique : ARNm constitué de régions codantes correspondant à plus d'un gène.

ARN nucléaire hétérogène (hn) : représente les transcrits des gènes nucléaires réalisés par l'ARN polymérase II ; il montre une large distribution de taille et une faible stabilité.

ARNsn (« small nuclear RNA » : petit ARN nucléaire) : n'importe laquelle des nombreuses espèces de petits ARN enfermés dans le noyau ; plusieurs ARNsn sont impliqués dans l'épissage ou d'autres réactions de maturation d'ARN.

ARNt correspondant : ARNt reconnu par une aminoacyl-ARNt synthétase particulière.

ARNt iso-accepteurs : ARNt acceptant le même acide aminé.

Arrêt de traduction par formation d'hybrides : technique qui permet l'identification de l'ADNc correspondant à un ARNm particulier en s'appuyant sur l'appariement de l'ARN *in vitro* pour inhiber la traduction.

Asque : l'asque d'un champignon contient une tétrade ou une octade de spores haploïdes qui représente les produits d'une méiose simple.

Assimilation de simple brin : décrit la capacité de la protéine RecA à faire déplacer l'un des brins d'une molécule d'ADN bicaténaire par un simple brin homologue ; le simple brin est alors assimilé dans le double brin.

Atténuateur : séquence de terminaison au niveau de laquelle s'effectue l'atténuation.

Atténuation : décrit la régulation de la terminaison de la transcription qui est impliquée dans le contrôle de l'expression de certains opérons bactériens.

Autoradiographie : détecte l'image créée sur un film photographique par des molécules marquées.

Autosome : tous les chromosomes exceptés les chromosomes sexuels ; une cellule diploïde a deux copies de chaque autosome.

Bactériophage : virus qui infecte des bactéries ; souvent désigné sous le terme de phage.

Bande de chromosome polytène : régions denses visibles qui contiennent la majorité de l'ADN ; les bandes de chromosomes normaux sont relativement plus grandes et sont engendrées sous forme de régions qui retiennent une coloration lors de certains traitements chimiques.

Bande majeure d'ADN génomique : consiste en un large pic sur un gradient de densité ; il n'est pas constitué des ADN satellites visibles qui, eux, forment des bandes séparées.

Bases modifiées : toutes les bases qui ne sont pas les quatre bases habituelles avec lesquelles l'ADN (T, C, A, G) ou l'ARN (U, C, A, G) sont synthétisés ; elles résultent de modifications post-synthétiques dans l'acide nucléique.

Bibliothèque : groupe de fragments clonés représentant toute la complexité de l'ADN cloné.

Bivalent : structure qui contient les 4 chromatides à chaque début de méiose (2 représentants de chaque homologue).

Blastoderme : étape de l'embryogenèse chez l'insecte dans laquelle une couche de noyaux ou de cellules autour de l'embryon entoure la masse interne du *vitellus*.

Boîte CAAT (Caat box) : fait partie d'une séquence conservée localisée environ 75 pb en amont du point de départ des unités de transcription eucaryotes. Est peut-être impliquée dans la liaison à l'ARN polymérase.

Boîte Hogness (TATA) : voir Boîte TATA.

Boîte homéotique : décrit la séquence conservée faisant partie de la région codante de *D. melanogaster* ; elle a été également trouvée dans les gènes d'amphibiens et de mammifères exprimés durant le développement embryonnaire précoce.

Boîte Pribnow (Pribnow box) : séquence consensus TATAATG centrée environ 10 pb avant le point de départ des gènes bactériens. Elle fait partie du promoteur spécialement important pour la liaison de l'ARN polymérase.

Boîte SOS : séquence d'ADN (opérateur) d'environ 20 pb qui est reconnue par la protéine de répression LexA.

Boîte TATA (de Hogness) (TATA box) : septamère conservé riche en A-T trouvé environ 25 pb avant le point de départ de l'unité de transcription de l'ARN polymérase II eucaryote ; elle pourrait être impliquée dans le positionnement de l'enzyme pour une initiation correcte.

Boucle : région simple brin à l'extrémité d'une épingle à cheveux dans l'ARN (ou l'ADN simple brin) ; elle correspond à la séquence se trouvant entre les séquences répétées inversées dans l'ADN double brin.

Boucle D : région d'ADN mitochondrial où un court fragment d'ARN est apparié à un brin d'ADN déplaçant ainsi le brin d'ADN d'origine, complémentaire de cette région. Le même terme est également utilisé pour décrire le déplacement d'une région d'un simple brin d'un ADN double brin par un composé simple brin envahisseur, au cours d'une réaction catalysée par la protéine RecA.

Boucle R : structure formée lorsqu'un brin d'ARN s'hybride avec un brin complémentaire dans un double brin d'ADN, déplaçant ainsi le brin initial ADN sous la forme d'une boucle s'étendant sur la région d'hybridation.

Brèche (gap) dans l'ADN : absence d'un ou plusieurs nucléotides sur un brin d'une molécule d'ADN double brin.

Brin anticodant : utilisé dans l'ADN double brin comme matrice pour diriger la synthèse de l'ARNt qui lui est complémentaire.

Brin codant : brin d'ADN qui a la même séquence que l'ARNm.

Brin principal (leading) d'ADN : brin d'ADN synthétisé de façon continue dans la direction 5'-3'.

Brin secondaire (lagging) d'ADN : brin d'ADN synthétisé dans la direction 3'-5' de façon discontinue sous la forme de courts fragments (5'-3') qui sont ensuite reliés entre eux de façon covalente.

Cadre de lecture : une des trois possibilités pour lire une séquence nucléotidique comme une série de triplets.

Cadre de lecture fermé : cadre contenant des codons de terminaison qui empêchent sa traduction en protéines.

CAP (CRP) : protéine de régulation positive d'*E. coli*, activée par l'AMP cyclique. Elle est nécessaire à l'ARN polymérase pour initier la transcription de certains opérons d'*E. coli* (sensibles aux catabolites).

Capside : manteau protéique externe d'une particule virale.

Carte de restriction : suite linéaire de sites sur l'ADN coupés par diverses enzymes de restriction.

Carte des territoires présomptifs (du destin) : carte d'un embryon qui montre les tissus adultes qui se développeront à partir des descendants des cellules qui occupent des régions particulières de l'embryon.

Caryotype : effectif chromosomique entier d'une cellule ou d'une espèce (visualisé durant la mitose).

Cassure et réunion : mode de recombinaison génétique par lequel deux molécules d'ADN double brin sont cassées en des points correspondants puis réunies après croisement.

C banding (technique des bandes C) : technique pour engendrer des régions colorées autour des centromères.

Cellule polyploïde : cellule possédant plus de deux fois un génome haploïde.

Cellules primaires : cellules eucaryotes mises en culture directement à partir de l'animal.

Cellules somatiques : toutes les cellules de l'organisme excepté celles de la lignée germinale.

Cellules T : lymphocytes de la lignée T (thymique) ; elles peuvent être sous-divisées en plusieurs types fonctionnels. Elles portent TcR (Récepteur des cellules T) et coopèrent avec les cellules B dans la création de la réponse immune.

Centromère : région compacte d'un chromosome qui comprend le site d'attachement au fuseau mitotique ou méiotique (voir aussi kinétochore).

Cercle roulant : mode de réplication dans lequel une fourche de réplication se meut le long d'une molécule circulaire pour un nombre indéfini de révolutions ; le brin d'ADN nouvellement synthétisé à chaque révolution déplace le brin synthétisé lors de la révolution précédente, donnant lieu à une queue contenant une série linéaire de séquences complémentaires du brin circulaire pris pour matrice.

Cercles d'ADN enchaînés : cercles d'ADN disposés comme des anneaux sur une chaîne.

Chiasma (pl chiasmata) : c'est ainsi que deux chromosomes homologues semblent avoir échangé du matériel pendant la méiose.

Chromatides : copies d'un chromosome produit par réplication. Le nom est habituellement utilisé pour les décrire dans la période qui précède leur séparation au cours de la division cellulaire en cours.

Chromatides sœurs : copies d'un chromosome produites lors de sa réplication.

Chromatine : complexe d'ADN et de protéine dans le noyau d'une cellule en interphase. Les chromosomes individuels ne peuvent y être distingués. Elle fut reconnue à l'origine par réaction avec des colorants spécifiques de l'ADN.

Chromocentre : agrégats d'hétérochromatine de différents chromosomes.

Chromomère : granule fortement coloré comme un constituant apparent du chromosome dans certaines conditions et particulièrement au début de la méiose.

Chromosome : unité distincte du génome portant beaucoup de gènes. Chaque chromosome est constitué d'une très longue molécule d'ADN double brin et d'une masse approximativement égale de protéines. Il n'est visible comme une entité morphologique qu'au cours de la division cellulaire.

Chromosome acrocentrique : chromosome dont le centromère est situé plus près d'une extrémité que de l'autre.

Chromosome à multiples fourches (dans la bactérie) : chromosome ayant plus d'une fourche de réplication, car un second site d'initiation est apparu avant que le premier cycle de réplication ait été terminé.

Chromosome dicentrique : produit de fusion de deux fragments chromosomiques dont chacun possède un centromère. Il est instable et peut être rompu quand les deux centromères sont attirés vers les pôles opposés pendant la mitose.

Chromosomes en écouvillons : grands chromosomes méiotiques trouvés dans les oocytes d'amphibiens.

Chromosomes polyténiques : chromosomes engendrés par réplications successives d'un groupe de chromosomes sans qu'il y ait séparation des répliques.

Chromosomes sexuels : chromosomes dont le contenu est différent entre les deux sexes ; ils sont généralement appelés X et Y (ou W et Z), un sexe a XX (ou WW), l'autre sexe a XY (ou WZ).

Cistron : unité génétique définie par le test *cis-trans* ; équivaut au gène en tant qu'unité d'ADN représentant une protéine.

Clone : décrit un grand nombre de cellules ou de molécules identiques issues d'une seule cellule ou molécule ancestrale.

Clone ADNc : séquence d'ADN double brin représentant un ARN porté par un vecteur de clonage.

Clones d'ADN génomique : séquences du génome portées par un vecteur de clonage.

Coconversion : correction simultanée de deux sites au cours de la conversion génique.

Code génétique : correspondance entre les triplets de l'ADN (ou l'ARN) et les acides aminés des protéines.

Codon : triplet de nucléotides qui représentent un acide aminé ou un signal de terminaison.

Codon ambre : le triplet nucléotidique UAG, un des trois codons non-sens qui provoque la terminaison de la synthèse protéique.

Codon de terminaison : un des trois triplets, UAG (ambre), UAA (ocre) et UGA qui provoque la terminaison de la synthèse protéique ; ils sont également appelés des codons non-sens.

Codon non-sens : un des trois triplets (UAG, UAA, UGA) qui provoque la fin de la synthèse protéique. (UAG est appelé ambre, UAA ocre.)

Codon ocre : triplet UAA, un des trois codons non-sens qui provoque la fin de la synthèse protéique.

Coévolution : voir évolution concertée.

Coiffe (cap) : structure à l'extrémité 5' de l'ARNm eucaryote introduite après la transcription, par liaison du phosphate terminal du 5'GTP à la base terminale de l'ARNm. Les G ajoutées (et quelquefois d'autres bases) sont méthylées donnant une structure de forme $7^{\text{Me}}\text{G}^5\text{ppp}^5\text{Np}...$

Complémentation : capacité de gènes indépendants (non alléliques) de fournir des produits diffusibles qui provoquent l'apparition du phénotype sauvage quand deux mutants sont testés en configuration *trans* chez un hétérozygote. Le test de complémentation *in vitro* consiste en l'identification d'un composant du type cellulaire sauvage qui peut conférer l'activité lorsqu'il est ajouté à un extrait préparé à partir de la cellule mutante. L'expérience identifie le composant rendu inactif par la mutation.

Complémentation interallélique : décrit un changement dans les propriétés d'une protéine multimérique occasionné par l'interaction de sous-unités codées par deux allèles mutants différents ; la protéine mixte peut être plus ou moins active que la protéine constituée des sous-unités de l'un ou l'autre type.

Complémentation négative : apparaît lorsque la complémentation interallélique permet à une sous-unité mutante de réprimer l'activité d'une sous-unité de type sauvage dans une protéine multimérique.

Complexe synaptonémal : décrit la structure morphologique de chromosomes en synapse.

Complexité : longueur totale de différentes séquences d'ADN présentes dans une préparation donnée.

Complexité chimique : quantité d'un composé ADN mesuré par une technique chimique.

Complexité cinétique : complexité d'un composant d'ADN mesurée par les cinétiques de réassociation d'ADN.

Commutation de classe : changement de l'expression de la région C d'une chaîne d'immunoglobuline au cours de la différenciation des lymphocytes.

Composant(s) intermédiaire(s) de la réaction de réassociation : composants qui réagissent entre les composants rapide (l'ADN satellite) et lent (l'ADN non répétitif) ; il est constitué d'ADN modérément répétitif.

Composant lent de la réaction de réassociation : c'est le dernier à se renaturer ; il contient de l'ADN hautement répétitif.

Composant rapide : dans une réaction de réassociation, c'est le premier à se renaturer ; il contient de l'ADN hautement répétitif.

Concatémère d'ADN : séries d'unités génomiques répétées en tandem.

Conditions non permissives : conditions ne permettant pas aux mutants létaux conditionnels de survivre.

Conditions permissives : conditions permettant aux mutants létaux conditionnels de survivre.

Configuration en *cis* : décrit deux sites sur la même molécule d'ADN.

Configuration en *trans* de deux sites : fait référence à leur présence sur deux molécules différentes d'ADN (chromosomes).

Conjugaison : décrit le « croisement » entre deux cellules bactériennes lorsque le chromosome (une partie) est transféré de l'une à l'autre.

Contrôle allostérique : se réfère à la capacité, qu'un site protéique peut avoir, d'influencer l'activité d'un autre site.

Contrôle autogène : décrit l'action du produit d'un gène qui, soit inhibe (contrôle autogène négatif), soit active (contrôle autogène positif) l'expression du gène par lequel il est codé.

Contrôle de réplication relâchée : fait référence à la capacité de certains plasmides de continuer leur réplication après que les bactéries aient cessé de se diviser.

Conversion génique : altération de l'un des brins d'un double brin hétérologue d'ADN le rendant complémentaire de l'autre au niveau où il y avait un mauvais appariement des bases.

Cordycépine : 3'déoxyadénosine, inhibiteur de la polyadénylation de l'ARN.

Corépresseur : petite molécule qui déclenche la répression de la transcription en se liant à une protéine régulatrice.

Correction sur épreuves : fait référence à tous les mécanismes qui corrigent les erreurs commises lors de la synthèse des protéines ou des acides nucléiques ; elle implique un examen

minutieux des unités individuelles après que celles-ci aient été incorporées dans la chaîne.

Cosmides : plasmides dans lesquels les sites *cos* du phage lambda ont été insérés ; il en résulte que l'ADN plasmidique peut être encapsidé *in vitro* dans le manteau du phage.

Cot : produit de la concentration d'ADN par le temps d'incubation au cours d'une réaction de réassociation.

Cot 1/2 : Cot requis pour atteindre la moitié de l'état final de la réaction d'association ; il est directement proportionnel à la longueur de l'ADN réassocié.

Cotransfection : transfection simultanée de deux marqueurs.

Coupures décalées dans un double brin d'ADN : coupures produites lorsque deux brins sont coupés à deux points différents mais proches l'un de l'autre.

Coupure (nick) dans un double brin d'ADN : absence d'une liaison phosphodiester entre deux nucléotides adjacents sur un brin.

Croisement en retour (back cross) : autre terme (plus ancien) pour test de croisement.

Crossing-over (croisement) : décrit l'échange réciproque de matériel qui se produit entre deux chromosomes pendant la méiose et est responsable de la recombinaison génétique.

Crossing-over inégaux : décrit un événement de recombinaison dans lequel les deux sites de recombinaison sont situés à des endroits différents (non alléliques) sur les deux molécules d'ADN parentales.

Cruciforme : structure produite dans l'ADN au niveau de répétitions inversées lorsque les séquences répétées s'appariaient avec leur complémentaire sur le même brin (à la place de l'appariement normal avec leur partenaire sur l'autre brin de la molécule double brin).

Cycle cellulaire : période s'écoulant entre deux divisions.

Cytosquelette : réseau de fibres dans le cytoplasme des cellules eucaryotes.

Dégénérescence : se réfère dans le code génétique à l'absence d'effet d'un changement quel qu'il soit, au niveau de la troisième base du codon, sur l'acide aminé qu'il représente.

Délétions : elles sont engendrées par l'enlèvement d'une séquence d'ADN, les régions situées de part et d'autre étant réunies.

Dénaturation : la dénaturation de l'ADN ou de l'ARN décrit sa conversion d'un état double brin à un état simple brin. La séparation des deux brins est le plus souvent accomplie par chauffage.

Dénaturation des protéines : décrit leur conversion d'une conformation physiologique en une autre conformation, inactive.

Densité de saturation : densité à laquelle les cellules eucaryotes en culture poussent *in vitro* avant que la division soit inhibée par les contacts inter-cellulaires.

Dépendance de l'ancrage : décrit le besoin, qu'ont les cellules eucaryotes normales, d'une surface à laquelle elles s'attachent pour croître en culture.

Dépendance du sérum : décrit le besoin des cellules eucaryotes en facteurs sériques afin de pouvoir pousser en culture.

Déplacement de brin : mode de réplication de certains virus dans lesquels un nouveau brin d'ADN grandit par déplacement du brin précédent (homologue) d'un double brin.

Descendance recombinante : descendance dont le génotype est différent de l'un ou l'autre des parents.

Développement précoce : se réfère à la période d'une infection phagique qui précède le début de la réplication de l'ADN, ou aux premiers stades de l'embryogenèse.

Dimère de thymine : paire de thymine adjacentes liées chimiquement dans l'ADN ; elle résulte de dommages induits par l'irradiation aux ultra-violets.

Diploïde : un ensemble diploïde de chromosomes contient deux copies de chaque autosome et deux chromosomes sexuels.

Disjonction : décrit le mouvement des membres d'une paire de chromosomes vers les pôles opposés durant la division cellulaire. A la mitose et au cours de la seconde division méiotique, la disjonction s'applique aux chromatides sœurs ; à la première division méiotique, elle s'applique aux paires de chromatides sœurs.

Distance sur la carte : mesurée en cM (centiMorgans) = le pourcentage de recombinaison (il est parfois soumis à des ajustements).

Divergence : pourcentage de différences entre les séquences nucléotidiques de deux ADN proches ou entre les séquences en acides aminés de deux protéines.

Diversification phénotypique : produite par une modification dans le génotype lors du développement somatique.

dna (mutants) : bactéries qui ne peuvent pas synthétiser l'ADN à 42 °C mais le font à 37 °C.

Domaine : le domaine d'un chromosome peut décrire soit une entité structurale distincte, définie comme une région où le superenroulement est indépendant des autres domaines, soit une région étendue contenant un gène exprimé dont la sensibilité à la dégradation par l'ADNase est augmentée.

Domaine d'une protéine : unité structurale composée d'une séquence d'acides aminés à laquelle une fonction particulière peut être attribuée.

Dominant (allèle) : détermine le phénotype d'un hétérozygote dont l'autre allèle est récessif.

Dosage génique : donne le nombre de copies d'un gène particulier dans le génome.

Down (mutation) : décrit les mutations d'un promoteur qui diminuent la fréquence d'initiation de la transcription.

Dysgénésie hybride : décrit l'incapacité de certaines lignées de *D. melanogaster* à se croiser entre elles, car les hybrides sont stériles (bien que par ailleurs ils soient phénotypiquement tout à fait normaux).

Échange simple brin : réaction dans laquelle une des chaînes d'un ADN double brin quitte son partenaire initial, et s'apparie à la chaîne complémentaire d'une autre molécule, déplaçant ainsi le simple brin homologue de cette seconde molécule d'ADN double brin.

Effet de position : fait référence au changement dans l'expression d'un gène dû à sa translocation à un nouveau site dans le génome ; par exemple, un gène actif peut devenir inactif s'il est placé près de l'hétérochromatine.

Élément de contrôle autonome : dans le maïs, transposon actif qui a la capacité de se transposer (cf. élément de contrôle non autonome).

Élément de contrôle du maïs : unité transposable à l'origine uniquement identifiée par ses propriétés génétiques. Ils peuvent être autonomes (capables de se transposer indépendamment) ou non autonomes (capables de se transposer uniquement en présence d'un élément autonome).

Éléments de contrôle non autonomes : transposons défectueux, qui ne peuvent se transposer que lorsqu'ils sont assistés par un élément de contrôle autonome de même type.

Empreinte protéique (finger print) : l'empreinte d'une protéine est le profil des fragments (habituellement obtenu par électrophorèse bidimensionnelle sur gel) engendré par clivage par une enzyme telle que la trypsine.

Empreinte sur l'ADN (foot printing) : technique d'identification d'un site ADN lié à une protéine grâce à la protection que fournit cette dernière contre la digestion des liaisons phosphodiester par la nucléase.

En amont : caractérise les séquences se trouvant dans la direction opposée de l'expression ; par exemple, le promoteur bactérien est en amont de l'unité de transcription, le codon d'initiation est en amont de la région codante.

Endonucléase : elle coupe les liaisons dans une chaîne d'acides nucléiques ; elles peuvent être spécifiques de l'ARN ou d'ADN simple brin ou double brin.

Endonucléase AP : coupe des sites apuriques ou apyrimidiques dans l'ADN du côté 5'.

Enroulement plectonémique : décrit l'entrecroisement des deux brins dans la double hélice d'ADN classique.

Enzymes de restriction : reconnaissent de courtes séquences spécifiques de l'ADN non méthylé (généralement) et coupent le double brin (parfois à un site cible, parfois autre part, selon le type).

Enzymes progressives : agissent en continu sur un substrat particulier, ce qui signifie qu'elles ne se dissocient pas entre chaque événement catalytique.

Épigénétiques (variations) : les variations épigénétiques influent sur le phénotype sans altérer le génotype.

Épingle à cheveux : décrit une région en double hélice, formée par l'appariement de bases entre des séquences complémentaires adjacentes dans un simple brin d'ADN ou d'ARN.

Épisome (plasmide) : matériel génétique extrachromosomique capable de s'intégrer dans l'ADN bactérien.

Épissage : décrit l'enlèvement des introns et la jonction des exons entre eux dans l'ARN ; ainsi, les introns sont éliminés alors que les exons sont épissés ensemble.

Épistatique : situation dans laquelle l'expression d'un gène annule les effets phénotypiques d'un autre gène.

Espacement de période courte : organisation dans le génome, dans laquelle des séquences d'ADN modérément répétitif d'environ 300 pb sont alternées avec des séquences d'ADN non répétitif d'environ 1 000 pb.

Espacement de période longue : organisation dans le génome, dans laquelle de longs segments d'ADN modérément répétitif et non répétitif sont alternés.

État déréprimé : décrit un gène qui est activé ; est synonyme d'état induit lorsqu'il décrit l'état normal d'un gène ; il a la même signification qu'état constitutif lorsqu'il décrit l'effet d'une mutation.

Eubactéries : comprennent la lignée majeure des procaryotes.

Euchromatine : comprend tout le génome d'un noyau à l'interphase, à l'exception de l'hétérochromatine.

Évolution coïncidente : voir évolution concertée.

Évolution concertée : décrit la capacité, qu'ont deux gènes proches, d'évoluer ensemble comme s'ils constituaient un locus unique.

Excision (d'un phage, d'épisomes ou d'autres séquences) : décrit sa libération du chromosome hôte sous forme d'une molécule.

Excision-réparation : mécanisme qui ôte une séquence simple brin d'un ADN qui contient des bases endommagées ou mal appariées, et qui la remplace dans le double brin en synthétisant une séquence complémentaire du brin restant.

Exclusion allélique : décrit l'expression, dans un lymphocyte donné, de l'un seulement des deux allèles codant pour l'immunoglobuline exprimée.

Exon : tout segment d'un gène interrompu qui est représenté dans l'ARN mature.

Exonucléase : coupe les nucléotides un par un, à partir de l'extrémité d'une chaîne polynucléotidique ; elle peut être spécifique soit de l'extrémité 5', soit de l'extrémité 3' de l'ARN ou de l'ADN.

Expérience de « playback » : décrit le recouvrement de l'ADN qui a été hybridé avec de l'ARN, et sa réutilisation dans une autre réaction de réassociation afin de savoir s'il est bien non répétitif.

Expériences de « pulse-chase » : sont réalisées en incubant très brièvement des cellules avec un précurseur marqué radioactivement de certaines voies métaboliques ou macromolécules ; la destinée du marquage est ensuite suivie durant une incubation avec un précurseur non marqué.

Expérience de « shotgun » (coup de fusil) : clonage d'un génome entier sous la forme de fragments engendrés au hasard.

Expérience d'hybridation par saturation : réalisée avec un large excès d'un composant, ce qui provoque l'entrée de toutes les séquences complémentaires d'un autre composant sous la forme double brin.

Extrémités cohésives : simples brins complémentaires d'ADN qui font saillie à partir des extrémités opposées d'un double brin ou d'extrémités de molécules double brin différentes ; elles peuvent être engendrées par des coupures décalées dans l'ADN double brin.

Facteur d'élongation (EF chez les procaryotes, eEF chez les eucaryotes) : protéines qui s'associent de façon cyclique avec les ribosomes permettant l'addition de chaque acide aminé à la chaîne polypeptidique.

Facteur F : sexe bactérien ou plasmide fertilisant.

Facteurs d'initiation (IF chez les procaryotes, eIF chez les eucaryotes) : protéines qui s'associent avec la petite sous-unité du ribosome spécifiquement à l'étape d'initiation de la synthèse protéique.

Facteurs de libération (de terminaison) : répondent aux codons non-sens pour provoquer la libération de la chaîne polypeptidique et du ribosome de l'ARNm.

Facteur rho : protéine qui assiste l'ARN polymérase de *E. coli* dans la terminaison de la transcription à certains sites (rho-dépendants).

Facteur sigma : sous-unité de l'ARN polymérase bactérienne qui est nécessaire à l'initiation ; il a l'influence principale sur la sélection des sites de liaison au niveau des promoteurs.

Famille Alu : ensemble de séquences apparentées, dispersées dans le génome humain et longues d'environ 300 bp. Chaque membre a des sites de clivage Alu aux extrémités (d'où leur nom).

Famille Alu-équivalente : ensemble de séquences appartenant à des génomes de mammifères et apparentées à la famille Alu humaine.

Famille de gènes : consiste en un groupe de gènes dont les exons sont apparentés ; les membres dérivent d'un gène ancestral par duplication et variation.

Figure en huit : décrit deux cercles d'ADN reliés par un événement de recombinaison qui n'est pas encore achevé.

Fixation par croisement (par cross-over) : se réfère aux conséquences possibles qui permettent à une mutation dans l'un des membres d'un groupe en tandem de s'étendre à tout le groupe (ou d'être limitée).

Formation de foyers : décrit la capacité des cellules eucaryotes transformées à pousser en groupes denses, empiétant les uns sur les autres.

Fourche de réplication : point à partir duquel les brins de l'ADN double brin parental sont séparés afin que la réplication puisse se faire.

Fragment acentrique d'un chromosome (engendré par cassure) : partie d'un chromosome qui ne contient pas le centromère et est perdue au cours de la division cellulaire.

Fragments d'Okasaki : courts segments de 1 000-2 000 bases produits lors de la réplication discontinue ; ils sont ensuite liés entre eux de façon covalente en un brin intact.

Fréquence de mutation : fréquence de mutants dans une population.

Fréquence de répétition : nombre (total) de copies d'une séquence donnée présentes dans un génome haploïde ; il est égal à 1 pour un ADN non répétitif, supérieur à 2 pour un ADN répétitif.

Fuseau : décrit la structure réorganisée d'une cellule eucaryote se divisant ; le noyau a été dissous et les chromosomes sont attachés au fuseau par des microtubules.

Fusion de l'ADN : signifie sa dénaturation.

Gamète : l'un ou l'autre des types de cellules reproductrices (germinales) — spermatozoïdes ou ovules — contenant un nombre haploïde de chromosomes.

G banding (technique de marquage des bandes G) : technique qui engendre des bandes dans les chromosomes en métaphase permettant de distinguer les membres d'un ensemble haploïde.

G1 : période du cycle cellulaire eucaryote qui s'étend de la dernière mitose au début de la réplication de l'ADN.

G2 : période du cycle cellulaire eucaryote qui s'étend de la fin de la réplication de l'ADN au début de la mitose suivante.

Gène (cistron) : segment d'ADN impliqué dans la production d'une chaîne polypeptidique ; il comprend des régions qui

précèdent ou qui suivent la région codante (séquences non codantes en 5' et séquences non codantes en 3') ainsi que les séquences intercalaires (introns) situées entre les séquences codantes (exons).

Gènes C : codent pour les régions constantes des chaînes d'immunoglobuline.

Gène constitutif : gène exprimé en fonction de l'interaction de l'ARN polymérase avec le promoteur, sans régulation supplémentaire ; ils sont parfois appelés gènes de ménage dans le contexte de la description de fonctions exprimées à bas niveau dans toutes les cellules.

Gènes de luxe : gènes codant pour des produits ayant des fonctions spécialisées, synthétisés (généralement en grandes quantités) dans des types cellulaires particuliers.

Gènes de ménage (constitutifs) : gènes qui sont (théoriquement) exprimés dans toutes les cellules, car ils remplissent les fonctions de base nécessaires à la survie de tous les types cellulaires.

Gènes de segmentation : gènes impliqués dans le contrôle du nombre ou de la polarité des segments du corps chez les insectes.

Gène de structure : code pour un ARN ou une protéine qui n'est pas régulateur(trice).

Gène extranucléaire : gène situé en dehors du noyau dans les organelles telles que mitochondries et chloroplastes.

Gènes homéotiques : gènes définis par des mutations qui convertissent une partie de l'organisme en une autre ; par exemple, une patte d'insecte peut être remplacée par une antenne.

Gène pléiotropique : affecte plus d'une caractéristique (apparemment non apparentées) du phénotype.

Gène régulateur : code pour un ARN ou une protéine dont la fonction est de contrôler l'expression d'autres gènes.

Gène V : séquence codant pour la majeure partie de la région variable (N-terminale) d'une chaîne d'immunoglobuline.

Génération F1 : c'est la première génération produite par croisement de deux lignées parentales.

Génotype : constitution génétique d'un organisme.

Groupe de compatibilité de plasmides : ensemble de plasmides incapables de coexister dans la même cellule bactérienne.

Groupe de complémentation : série de mutations incapables de compléter quand elles sont testées par paires en *trans* ; définit une unité génétique (le cistron) qui devrait plutôt être appelé un groupe de non-complémentation.

Groupe de gènes : groupe de gènes adjacents qui sont identiques ou apparentés.

Groupe de liaison (de linkage) : comprend tous les loci qui peuvent être apparentés par des relations de liaison (directement ou indirectement) ; il est équivalent à un chromosome.

Gyrase : topo-isomérase de type II de *E. coli* qui a la capacité d'introduire des supertours négatifs dans l'ADN.

Haploïde : un groupe de chromosomes haploïdes est constitué d'une copie de chaque autosome et d'un chromosome sexuel ; le nombre haploïde n est caractéristique des gamètes des organismes diploïdes.

Haptène : petite molécule qui agit comme un antigène quand elle est conjuguée à une protéine.

Hémizygote : individu diploïde qui a perdu une copie d'un gène particulier (parce qu'un chromosome a été perdu, par exemple) et qui, par conséquent, n'a plus qu'une seule copie de ce gène. Chez les mammifères, les mâles sont obligatoirement hémizygotes pour les gènes localisés sur l'X.

Hérédité cytoplasmique : propriété des gènes localisés dans les mitochondries ou les chloroplastes (ou peut-être dans d'autres organelles extra-nucléaires).

Hérédité maternelle : décrit la survie préférentielle dans la descendance, de marqueurs génétiques apportés par la mère.

Hétérochromatine : décrit des régions du génome qui sont hautement condensées en permanence et qui ne sont pas exprimées génétiquement. Elle peut être constitutive ou facultative.

Hétérochromatine constitutive : décrit l'état inerte de séquences en permanence non exprimées, habituellement ADN satellite souvent localisé au niveau du centromère.

Hétérochromatine facultative : décrit l'état inerte de séquences qui existent aussi sous forme de copies actives — par exemple, un chromosome X de mammifère chez les femelles.

Hétérozygote : individu qui possède des allèles différents à des loci particuliers.

Histones : protéines conservées se liant à l'ADN chez les eucaryotes et qui forment le nucléosome, sous-unité de base de la chromatine.

Homologues : chromosomes portant les mêmes loci génétiques ; une cellule diploïde a deux copies de chaque homologue, chacune dérivée d'un des parents.

Homozygote : individu qui a le même allèle aux mêmes loci sur des chromosomes homologues.

Horloge de l'évolution : elle est définie par la vitesse à laquelle les mutations s'accumulent dans un gène donné.

Hybridation : appariement entre des chaînes complémentaires d'ADN ou d'ARN.

Hybridation cytologique : voir hybridation *in situ*.

Hybridation en milieu liquide : réaction d'hybridation réalisée en solution entre des brins d'acides nucléiques complémentaires.

Hybridation en solution : voir hybridation en milieu liquide.

Hybridation guidée par l'ADN : implique la réaction d'un excès d'ADN avec l'ARN.

Hybridation *in situ* : réalisée en dénaturant l'ADN des cellules écrasées sur une lame de microscope de telle façon que la réaction soit possible en ajoutant de l'ARN ou de l'ADN simple brin ; la préparation ajoutée est marquée radioactivement et son hybridation est suivie par une autoradiographie. Elle peut également être utilisée pour détecter des ARN cellulaires.

Hybridation sur colonies : technique utilisant l'hybridation *in situ* pour identifier des bactéries portant des vecteurs chimériques dont l'ADN inséré est homologue d'une séquence particulière.

Hybridation sur filtre : elle s'accomplit en incubant une préparation d'ADN dénaturée ou d'ARN, immobilisée sur un filtre de nitro-cellulose avec une solution d'ADN ou d'ARN marquée.

Hybridome : lignée cellulaire produite par la fusion d'un myélocyte et d'un lymphocyte ; il continue indéfiniment à exprimer l'immunoglobuline du ou des parents sécréteurs.

Hyperchromicité : augmentation de la densité optique qui se produit quand l'ADN est dénaturé.

Hypothèse du signal : décrit le rôle de la séquence N-terminale d'une protéine sécrétée, dans l'attachement d'un polypeptide naissant à la membrane ; elle signifie que l'ARNm et le ribosome sont attachés à la membrane par l'extrémité N-terminale de la protéine en train d'être synthétisée.

Hypothèse de Wobble (oscillation) : rend compte de la capacité d'un ARNt à reconnaître plus d'un codon par l'appariement inhabituel (non G-C, A-T) de la troisième base d'un codon.

Immortalisation : décrit l'acquisition par une lignée cellulaire eucaryote de la capacité à croître et à se diviser indéfiniment en culture.

Immunité des phages : fait référence à la capacité d'un prophage à empêcher un autre phage du même type d'infecter une cellule. Elle résulte de la synthèse d'un répresseur du phage par le génome du prophage.

Immunité des plasmides : décrit la capacité d'un plasmide à empêcher un autre plasmide du même type de s'établir dans une cellule. Elle résulte généralement d'une interférence avec l'aptitude à se répliquer de ce plasmide.

Immunité des transposons (ou de transposition) : fait référence à la capacité de certains transposons à empêcher d'autres transposons du même type de se transposer sur la même molécule d'ADN. Elle résulte d'une variété de mécanismes.

Immunité lysogénique : capacité d'un prophage à empêcher un autre génome de phage du même type de s'établir dans la bactérie.

Incompatibilité : incapacité de certains plasmides bactériens de coexister dans la même cellule. Elle est la cause de l'immunité des plasmides.

Indice de flottaison : mesure la capacité d'une substance à flotter dans un fluide standard, par exemple le chlorure de césium (CsCl).

Inducteur : petite molécule qui déclenche la transcription d'un gène en se liant à une protéine régulatrice.

Inducteurs gratuits : ressemblent aux vrais inducteurs de la transcription, mais ne sont pas des substrats pour les enzymes induites.

Induction : fait référence à la capacité d'une bactérie (ou d'une levure) de synthétiser certaines enzymes, uniquement lorsque leurs substrats sont présents ; appliquée à l'expression des gènes, elle fait référence au déclenchement de la transcription grâce à l'interaction de l'inducteur avec la protéine régulatrice.

Induction d'un prophage : décrit son excision du génome de l'hôte et son entrée dans un cycle lytique (infectieux), résultat de la destruction du répresseur lysogénique.

Infection lytique d'une bactérie par un phage : infection qui conduit à la destruction des bactéries et à la libération de la descendance du phage.

Inhibition par le produit : capacité d'un produit d'une voie métabolique d'inhiber une enzyme qui catalyse une des premières étapes de cette voie.

Insertions : identifiées par la présence d'un segment de paires de bases supplémentaires dans l'ADN.

Intégration d'une séquence virale ou autre séquence d'ADN : insertion dans le génome de l'hôte d'une séquence liée de façon covalente à l'ADN de l'hôte.

Interbandes : régions relativement dispersées des chromosomes polytènes qui se trouvent entre les bandes.

Interphase : période entre les divisions cellulaires mitotiques ; elle est divisée en G1, S et G2.

Intron : segment d'ADN qui est transcrit, mais excisé du transcrit, alors que sont épissés les exons situés de part et d'autre.

Inversion : modification chromosomique dans laquelle un segment a été pivoté de 180° par rapport aux régions se trouvant de chaque côté, puis réinséré.

IS : abréviation pour séquence d'insertion (« insertion sequence »), un petit transposon bactérien portant uniquement les fonctions génétiques impliquées dans la transposition.

Isomères topologiques : molécules d'ADN identiques, excepté en ce qui concerne leur « linking number ».

Isotype : groupe de chaînes d'immunoglobulines étroitement apparentées.

Jonctions d'épissage : séquences entourant immédiatement les liaisons exon-intron.

Jonction d'épissage droite : liaison entre l'extrémité droite d'un intron et l'extrémité gauche de l'exon adjacent.

Jonction d'épissage gauche : liaison entre l'extrémité droite d'un exon et l'extrémité gauche d'un intron.

Jonction de recombinaison : point reliant deux molécules recombinantes d'ADN double brin (bord de la région double brin hétérologue).

Jonction paranémique : décrit une région dans laquelle deux séquences complémentaires d'ADN sont associées côte à côte, au lieu d'être entrecroisées dans une structure en double hélice.

kb : abréviation pour 1 000 paires de bases d'ADN ou 1 000 bases d'ARN.

Kinétocore : structure du chromosome à laquelle s'attachent les microtubules du fuseau mitotique (voir également centromère).

Leader : séquence non traduite à l'extrémité 5' de l'ARN_m qui précède le codon d'initiation.

Liaison (linkage) : décrit la tendance des gènes à être hérités ensemble à cause de leur localisation sur le même chromosome : elle est mesurée par le pourcentage de recombinaison entre loci.

Liaison (ou ligature) à bouts francs (blunt end) : réaction qui unit deux molécules d'ADN double brin directement à leurs extrémités.

Liaison au sexe : modèle d'hérédité des gènes portés sur un chromosome sexuel (généralement l'X).

Liaison-ligature (ligation) : formation d'une liaison phosphodiester pour lier deux bases adjacentes séparées par une coupure sur l'un des brins de la double hélice de l'ADN. (Ce terme peut également s'appliquer à la liaison d'extrémités franches et à la jonction d'ARN.)

Lignées cellulaires établies : lignées constituées de cellules eucaryotes immortalisées.

Lignées petites de levure : lignées de levure qui ont perdu la fonction mitochondriale.

Linker : court oligonucléotide double brin synthétique comprenant le site cible d'une enzyme de restriction ; il peut être ajouté aux extrémités d'un fragment d'ADN coupé par une autre enzyme de restriction durant la reconstruction de l'ADN recombinant.

Linking number (nombre de liaison) : nombre de fois où les deux brins d'un double brin d'ADN fermé se croisent l'un l'autre.

Locus : position sur un chromosome où un gène particulier réside ; le locus peut être occupé par n'importe quel allèle de ce gène.

Locus agissant en *cis* : affecte uniquement l'activité de séquences d'ADN situées sur sa propre molécule d'ADN. Cette propriété implique habituellement que le locus ne code pas pour une protéine.

Locus complexe (de *D. melanogaster*) : locus ayant des propriétés génétiques incompatibles avec la fonction d'un gène représentant une seule protéine. Les loci complexes sont habituellement très grands, supérieurs à 100 kb au niveau moléculaire.

Locus létal : n'importe quel gène dans lequel une mutation létale peut être obtenue.

Locus majeur d'histocompatibilité : grande région chromosomique constituée d'un groupe géant de gènes qui codent pour les antigènes de transplantation et d'autres protéines trouvées à la surface des lymphocytes.

LTR : abréviation pour « long terminal repeat » (longue séquence répétée terminale), une séquence directement répétée aux deux extrémités d'un ADN rétroviral.

Lymphocytes B : ou cellules B, cellules responsables de la synthèse des anticorps.

Lyse : décrit la mort de bactéries à la fin d'un cycle infectieux phagique lorsque celles-ci éclatent, relâchant la descendance du phage infectieux.

Lysogène : caractérise une bactérie lorsqu'elle possède un prophage réprimé dans son génome.

Lysogénie : décrit la capacité d'un phage à survivre dans une bactérie comme un composant prophagique stable du génome bactérien.

Marche sur le chromosome (chromosome walking) : décrit l'isolement séquentiel de clones portant des séquences d'ADN qui se chevauchent, permettant à de grandes régions du chromosome d'être explorées. La marche est souvent accomplie afin d'atteindre un locus particulier.

Marquage d'extrémité : décrit l'addition d'un groupe marqué à une extrémité (5' ou 3') d'un brin d'ADN.

Marquage indirect des extrémités : technique qui permet d'examiner l'organisation de l'ADN en coupant à un site spécifique et en isolant tous les fragments contenant la séquence adjacente à un côté de la coupure ; il révèle la distance entre la coupure et la (les) prochaine(s) cassure(s) dans l'ADN.

Marqueur (ADN) : fragment de taille connue utilisé pour calibrer un gel d'électrophorèse.

Marqueur génétique : tout allèle étudié dans une expérience.

Matrice nucléaire : réseau de fibres entourant et pénétrant le noyau.

Méiose : se produit par deux divisions successives (méiose I et méiose II) qui réduisent le nombre de départs de 4n chromosomes à 1n dans chacune des quatre cellules produites. Les cellules peuvent ensuite poursuivre leur maturation en cellules germinales (spermatozoïdes ou ovules).

Métastase : décrit la capacité des cellules tumorales à quitter leur site d'origine et à migrer à d'autres endroits du corps, où une nouvelle colonie est établie.

Migration de la branche : décrit la capacité d'un brin d'ADN partiellement apparié avec son complémentaire d'étendre son appariement en déplaçant le brin résident dont il est homologue.

Minichromosome de SV40 ou de polyome : forme nucléosomale de l'ADN viral circulaire.

Mitose : division d'une cellule somatique eucaryote.

Modification de l'ADN ou de l'ARN : comprend toutes les modifications effectuées sur les nucléotides après leur incorporation initiale dans la chaîne polynucléotidique.

Monocouche : décrit la croissance de cellules eucaryotes en culture en une couche épaisse d'une seule cellule.

Mutagènes : augmentent le taux de mutation en induisant des changements dans l'ADN.

Mutants *dna* de *E. coli* à arrêt rapide : cessent immédiatement leur réplication quand la température est élevée à 42 °C.

Mutants *dna* de *E. coli* à arrêt lent : terminent le cycle de la réplication bactérienne en cours, mais ne peuvent en initier un autre à 42 °C.

Mutants à lyse rapide (r) : montrent une modification dans le mécanisme de lyse de *E. coli* à la fin d'une infection par un phage T-even (pair).

Mutants non inductibles : ne peuvent être induits.

Mutants phagiques virulents : incapables d'établir une lysogénie.

Mutants relâchés de *E. coli* : n'ont pas de réponse forte à la privation en acides aminés (ou autres privations nutritionnelles).

Mutant sensible au froid : mutant défectif à basse température mais fonctionnel à température normale.

Mutation : décrit tout changement dans la séquence de l'ADN génomique.

Mutation ambre : décrit tout changement dans l'ADN qui crée un codon ambre dans un site occupé auparavant par un codon représentant un acide aminé dans une protéine.

Mutation constitutive : entraîne l'expression sans régulation de gènes habituellement régulés.

Mutations directes (forward) : inactivent un gène sauvage.

Mutations du cadre de lecture (frame shift) : apparaissent par délétion ou insertion de séquences qui ne sont pas des

multiples de 3 pb ; elles changent le cadre dans lequel les triplets sont traduits en protéine.

Mutations induites : résultent de l'addition d'un mutagène.

Mutations « leaky » : permettent un taux d'expression des gènes résiduel.

Mutations létales conditionnelles : entraînent la mort d'une cellule ou d'un virus dans certaines conditions (non permissives), mais permettent leur survie dans d'autres conditions (permissives).

Mutation non-sens : toute modification dans l'ADN qui donne lieu à un codon non-sens (de terminaison) en remplacement d'un codon représentant un acide aminé.

Mutation ocre : toute modification dans l'ADN qui crée un codon UAA à un site occupé précédemment par un autre codon.

Mutations par balayage de linker (linker scanning mutation) : elles sont introduites en recombinant deux molécules d'ADN *in vitro* à un fragment de restriction ajouté à l'extrémité de chacune, ce afin d'insérer la séquence linker à un site de recombinaison.

Mutation ponctuelle : modification impliquant une seule paire de bases.

Mutation reverse : annule l'effet d'une mutation qui avait inactivé un gène ; le type sauvage est ainsi restauré.

Mutations silencieuses : ne modifient pas le produit du gène.

Mutation somatique : mutation se produisant dans une cellule somatique, et n'affectant, par conséquent, que ses descendants ; elle n'est pas héréditaire.

Mutations spontanées : se produisent en l'absence de tout réactif ajouté pour augmenter le taux de mutation.

Mutation thermo-sensible : crée un produit génique qui est fonctionnel à basse température, mais inactif à température élevée (la relation inverse est généralement appelée mutation sensible au froid).

Mutations « up » du promoteur : augmentent la fréquence d'initiation de la transcription.

Myélome : lignée de cellules tumorales dérivée d'un lymphocyte ; il produit généralement un seul type d'immunoglobuline.

Nick translation (déplacement de coupure) : décrit la capacité de l'ADN polymérase I de *E. coli* d'utiliser une coupure comme point de départ pour dégrader un des brins de l'ADN double brin et de le remplacer par synthèse de nouveau matériel ; elle est utilisée pour introduire des nucléotides marqués radioactivement dans l'ADN *in vitro*.

Nodules de recombinaison (« nodes ») : structures denses présentes dans le complexe synaptonémal ; elles peuvent être impliquées dans les crossing-over.

Nombre de torsions d'un ADN (twisting number) : nombre de paires de bases divisé par le nombre de paires de bases par tour de la double hélice.

Non-disjonction : décrit l'incapacité des chromatides (chromosomes dupliqués) à migrer vers les pôles opposés au cours de la mitose ou de la méiose.

Northern blotting : technique pour transférer l'ARN d'un gel d'agarose à un filtre de nitrocellulose sur lequel il peut être hybridé à de l'ADN complémentaire.

Nucléase micrococcale : endonucléase qui coupe l'ADN ; dans la chromatine, l'ADN est clivé préférentiellement entre les nucléosomes.

Nucléase S1 : enzyme qui dégrade spécifiquement les séquences d'ADN non appariées (simple brin).

Nucléoïde : corps compact qui renferme le génome dans une bactérie.

Nucléole : région définie du noyau créée par la transcription des gènes codant pour les ARNr.

Nucléosome : sous-unité structurale de base de la chromatine, elle consiste en environ 200 pb d'ADN et en un octamère de protéines histones.

Œil de réplication : région dans laquelle l'ADN a été répliqué à l'intérieur d'une région plus longue non répliquée.

Oncogènes : gènes dont les produits ont la capacité de transformer des cellules eucaryotes de telle sorte qu'elles poussent de la même façon que des cellules tumorales. Les oncogènes portés par des rétrovirus ont des noms sous la forme *v-onc*. Voir également proto-oncogènes.

Opérateur : site sur un ADN procaryotique auquel une protéine de répression se lie afin d'empêcher la transcription d'être initiée au promoteur adjacent.

Opéron : unité d'expression et de régulation de gènes bactériens ; elle est constituée de gènes de structure et d'éléments de contrôle dans l'ADN reconnus par un (des) produit(s) de gène(s) régulateur(s).

Organisateur nucléolaire : région d'un chromosome portant des gènes codant pour des ARNr.

Organismes procaryotes (bactéries) : n'ont pas de noyau.

Organismes « true breeding » : organismes homozygotes pour le trait considéré.

Origine (ori) : séquence d'ADN où la réplication est initiée.

Orphons : gènes individuels isolés et trouvés dans des régions isolées, mais apparentés à des membres d'un groupe de gènes.

Paires de bases (pb) : appariements entre A et T et G et C dans un ADN double brin ; d'autres paires peuvent être formées dans l'ARN, dans certaines conditions.

Palindrome : séquence d'ADN qui est la même que l'on lise un brin de la gauche vers la droite, ou l'autre brin de la droite vers la gauche ; il consiste en des séquences répétées inversées adjacentes.

Papovavirus : représentent une classe de virus animaux à petit génome, elle comprend SV40 et polyome.

Particule cœur : produit de digestion du nucléosome qui retient l'octamère d'histones et comporte 146 pb d'ADN ; sa structure apparaît similaire à celle du nucléosome lui-même.

pb : abréviation pour paire de bases ; la distance le long de l'ADN est mesurée en pb.

pBR322 : un des vecteurs de clonage plasmidiques usuels.

Période tardive du développement des phages : période d'infection qui suit le début de la réplication de l'ADN.

Périodicité de l'ADN : nombre de paires de bases par tour de la double hélice.

Phage (bactériophage) : virus bactérien.

Phase ouverte de lecture (ORF : « open reading frame ») : est constituée d'un série de triplets codant pour des acides aminés sans aucun codon de terminaison ; la séquence peut être (potentiellement) traduite en une protéine.

Phase S : partie restreinte du cycle cellulaire eucaryotique durant laquelle la synthèse de l'ADN se produit.

Phénotype : apparence (ou autres caractéristiques) d'un organisme, résultant de l'interaction de sa constitution génétique avec l'environnement.

Plasmide : ADN circulaire extrachromosomique autonome autorépliatif.

Plasmides à copie unique : se maintiennent dans les bactéries à un taux de un plasmide par chromosome hôte.

Plasmides multi-copies : sont présents dans les bactéries à des taux plus importants qu'un par chromosome.

Plasmide sexuel : est en réalité un épisome ; il est capable d'initier le processus de conjugaison, par lequel du matériel chromosomique est transféré d'une bactérie à l'autre.

Ploïdie : se réfère au nombre de copies du groupe de chromosomes présents dans une cellule ; un haploïde a une copie, un diploïde a deux copies, etc.

Point chaud (hot spot) : site au niveau duquel la fréquence de mutation (ou de recombinaison) est très augmentée.

Point de départ, site d'initiation (site de départ) : fait référence à la position sur l'ADN correspondant à la première base incorporée dans l'ARN.

Polarité : se réfère à l'effet d'une mutation dans un gène, qui se manifeste en influençant l'expression (transcription ou tra-

duction) de gènes consécutifs dans la même unité de transcription.

Polyadénylation : addition d'une séquence d'acide polyadénylique à l'extrémité 3' d'un ARN eucaryote après sa transcription.

Polymorphisme : se réfère à l'apparition simultanée dans une population, de génomes montrant des variations alléliques (dans les allèles produisant différents phénotypes ou, par exemple, dans les modifications de l'ADN affectant le profil de restriction).

Polyprotéine : produit d'un gène qui est clivé en plusieurs protéines indépendantes.

Polysome (polyribosome) : ARNm associé à une série de ribosomes engagés dans la traduction.

Primosome : décrit le complexe de protéines impliquées dans l'amorçage qui initie la synthèse de chaque fragment d'Okasaki lors de la réplication discontinue de l'ADN ; le primosome peut se déplacer le long de l'ADN, afin de s'engager dans des événements d'amorçage successifs.

Promoteur : région de l'ADN impliquée dans la liaison de l'ARN polymérase à l'ADN lors de l'initiation de la transcription.

Prophage : génome phagique intégré de façon covalente comme une partie linéaire du chromosome bactérien.

Protéine agissant en cis : a la propriété exceptionnelle d'agir seulement sur la molécule d'ADN à partir de laquelle elle est exprimée.

Protéines d'antiterminaison : permettent à l'ARN polymérase de transcrire au-delà de certains sites terminateurs.

Protéines de régulation positive : sont requises pour l'activation d'une unité de transcription.

Protéine de répression : se lie à l'opérateur sur l'ADN ou sur l'ARN afin d'empêcher respectivement la transcription ou la traduction.

Protéines hétéromultimériques : sont constituées de sous-unités différentes (codées par des gènes différents).

Protéine homomultimérique : protéine constituée de sous-unités identiques.

Protéines multimériques : sont constituées de plus d'une sous-unité.

Proto-oncogènes : équivalents normaux dans le génome eucaryote des oncogènes portés par certains rétrovirus. On leur donne des noms sous la forme *c-onc*.

Provirus : séquence d'ADN double brin dans le chromosome eucaryote qui correspond au génome d'un rétrovirus à ARN.

Pseudogènes : composants inactifs mais stables du génome, dérivés par mutation d'un gène ancestral actif.

Puff (nodule) : expansion d'une bande de chromosome polytène qui est associée à la synthèse d'ARN à certains loci de la bande.

Queue (trailer) : séquence non traduite à l'extrémité 3' d'un ARNm après le codon de terminaison.

Rapport de condensation : rapport de la longueur d'ADN sur l'unité de longueur de la fibre le contenant.

Réaction « au ralenti » (idling) : production de pppGpp et de ppGpp par des ribosomes lorsqu'un ARNt non chargé est présent à un site A ; elle déclenche une réaction stringente.

Réactions d'hybridations conduites par l'ARN : utilisent un excès d'ARN pour réagir avec toutes les séquences complémentaires dans une préparation d'ADN simple brin.

Réassociation de l'ADN : décrit l'appariement de simples brins complémentaires pour former une double hélice.

RecA : produit du locus *recA* de *E. coli* ; c'est une protéine à activité double, elle agit comme une protéase et est aussi capable d'échanger des simples brins de molécules d'ADN. L'activité protéasique contrôle la réponse SOS ; la capacité de manipulation des acides nucléiques est impliquée dans les voies de la recombinaison-réparation.

Recombinaison conservative : implique la cassure et la réunion de brins d'ADN préexistants sans aucune synthèse de nouvelles séquences d'ADN.

Recombinaison réciproque : production de nouveaux génotypes avec des arrangements inverses d'allèles par rapport à l'origine maternelle ou paternelle.

Recombinaison-réparation : moyen de remplir une brèche dans un brin d'un double brin d'ADN, grâce à l'utilisation d'un simple brin homologue provenant d'un autre double brin.

Recombinaison spécifique de sites : se produit entre deux séquences spécifiques (pas nécessairement homologues), lors de l'intégration/excision d'un phage ou de la résolution de structures co-intégrées lors de la transposition.

Redondance terminale : décrit la répétition de la même séquence aux deux extrémités d'un génome phagique (par exemple).

Régions constantes : régions des immunoglobulines codées par les gènes C qui font partie de la chaîne qui varie le moins et identifie le type de l'immunoglobuline.

Régions hypervariables d'une immunoglobuline : parties de la région variable qui montrent la variabilité maximale lorsque l'on compare différents anticorps.

Région intercistronique : distance entre le codon de terminaison d'un gène et le codon d'initiation du gène suivant.

Région variable d'une chaîne d'immunoglobuline : est codée par le gène V et est extrêmement variable lorsque l'on compare différentes chaînes d'immunoglobulines ; elle provient de copies génomiques multiples (différentes) et des modifications introduites lors de la construction d'une immunoglobuline active.

Règle GT-AG : décrit la présence constante de ces nucléotides aux deux premières et deux dernières positions des introns des gènes nucléaires.

Régulateurs négatifs : fonctionnent en éteignant la transcription ou la traduction.

Régulation coordonnée : se réfère au contrôle commun d'un groupe de gènes.

Renaturation : réassociation de simples brins complémentaires dénaturés d'une double hélice d'ADN. Elle est également utilisée pour décrire le rétablissement de la structure d'une protéine dénaturée.

Répétitions directes : séquences identiques (ou apparentées) présentes en deux ou plus de deux copies dans la même orientation et dans la même molécule d'ADN ; elles ne sont pas nécessairement adjacentes.

Réplication bidirectionnelle : s'accomplit quand deux fourches de réplication se déplacent à partir de la même origine dans des directions différentes.

Réplication discontinue : se réfère à la synthèse d'ADN sous forme de courts fragments (Okasaki) qui sont reliés plus tard en un brin continu.

Réplication saltatoire : amplification latérale soudaine qui produit un grand nombre de copies d'une séquence particulière.

Réplication semi-conservative : est réalisée par séparation des brins du double brin d'ADN parental, chacun d'eux agissant ensuite comme matrice pour la synthèse d'un brin complémentaire.

Réplication semi-discontinue : mode de réplication selon lequel un nouveau brin est synthétisé de façon continue, alors que l'autre est synthétisé de façon discontinue.

Réplication stringente : décrit la limitation des plasmides à copie unique à ne se répliquer que parallèlement au chromosome bactérien.

Réplication uni-directionnelle : fait référence au mouvement d'une seule fourche de réplication à partir d'une origine donnée.

Réplicon : unité génomique de réplication de l'ADN ; il contient une origine pour l'initiation de la réplication.

Réplisome : structure multiprotéique qui est assemblée à l'endroit de la fourche de réplication bactérienne afin d'entreprendre la synthèse de l'ADN. Il contient l'ADN polymérase et d'autres enzymes.

Réponse SOS chez *E. coli* : décrit l'induction coordonnée de nombreuses enzymes, dont celles ayant une activité de réparation, en réponse à l'irradiation ou autre dommage subi par l'ADN ; elle résulte de l'activation de la protéase RecA pour cliver le répresseur LexA.

Réponse stringente : se réfère à la capacité d'une bactérie d'arrêter la synthèse d'ARNt et de ribosomes dans un milieu de culture pauvre.

Répresseur lysogénique : protéine qui empêche un prophage d'entrer à nouveau dans un cycle lytique.

Répression : capacité des bactéries d'empêcher la synthèse de certaines enzymes quand leurs produits sont présents ; plus généralement, elle fait référence à l'inhibition de la transcription (ou de la traduction) par la liaison d'une protéine de répression à un site spécifique sur l'ADN (ou l'ARNm).

Répression catabolique : décrit la diminution de l'expression de beaucoup d'opérons bactériens qui résulte de l'addition de glucose ; elle est causée par une diminution du niveau de l'AMPc qui, à son tour, inactive le régulateur CAP.

Résolvase : activité enzymatique impliquée dans la recombinaison spécifique de sites entre deux transposons présents en tant que séquences répétées directes dans une structure co-intégrée.

Rétro-pseudogène ou pseudogène ayant subi une maturation (processed gene) : copie inactive d'un gène qui n'a pas d'introns, elle contraste avec la structure morcelée du gène actif. Ces gènes tiennent probablement leur origine de la transcription inverse de l'ARNm puis de l'insertion de la copie double brin dans le génome.

Rétro-régulation : décrit la capacité d'une séquence dans le transcrit située en aval de la région codante, à réguler la traduction d'un ARNm.

Rétrovirus : virus à ARN qui se propage par conversion en ADN double brin.

Réversion d'une mutation : changement dans l'ADN qui soit inverse l'altération initiale (réversion vraie), soit la compense (réversion d'un second site dans le même gène).

Révertants : dérivés par réversion d'une cellule ou d'un organisme mutant.

Rifamycines (incluant la rifampicine) : inhibent la transcription dans les bactéries.

Rot : produit de la concentration d'ARN et du temps d'incubation dans une réaction d'hybridation conduite par l'ARN.

Satellite cryptique : séquence d'ADN satellite non identifiée en tant que telle comme faisant partie d'un pic séparé en gradient de densité ; c'est-à-dire qu'elle reste présente dans la bande d'ADN principale.

Scyrps : ribonucléoprotéines cytoplasmiques (ARNsc associés à des protéines).

Ségrégation post-méiotique : décrit la ségrégation de deux brins d'un double brin d'ADN qui portent chacun une information différente (créée par la formation d'un double brin hétérologue lors de la méiose) lorsque la réplication suivante permet aux brins de se séparer.

Sélection : décrit l'utilisation de certaines conditions pour permettre uniquement la survie de cellules d'un phénotype particulier.

Séquences aval : séquences s'étendant en aval par rapport à la direction de l'expression ; par exemple, la région codante est en aval du codon d'initiation.

Séquence *chi* : octamère qui fournit un point chaud pour la recombinaison génétique dépendante de RecA chez *E. coli*.

Séquence consensus : séquence idéalisée dans laquelle chaque position représente la base la plus souvent trouvée quand de nombreuses séquences connues sont comparées.

Séquence de Shine-Dalgarno : partie ou totalité de la séquence polypurinique AGGAGG située sur l'ARNm bactérien juste avant son codon d'initiation AUG ; elle est complémentaire de la séquence se trouvant à l'extrémité 3' de l'ARNr 16S ; elle est impliquée dans la liaison du ribosome à l'ARNm.

Séquence d'insertion (IS) : petit transposon bactérien qui porte uniquement les gènes nécessaires à sa propre transposition.

Séquence intercalaire : intron.

Séquence intercalaire non transcrite : région se trouvant entre les unités de transcription dans un groupe de gènes en tandem.

Séquence intercalaire transcrite : partie d'une unité de transcription d'un ARNr qui est transcrite, mais éliminée lors de la maturation ; ce qui signifie qu'on ne la retrouve pas dans l'ARNr final.

Séquences répétées en tandem : copies multiples de la même séquence, en série.

Séquences répétées inversées : sont constituées de deux copies de la même séquence d'ADN répétée mais en orientation opposée sur la même molécule. Des séquences répétées inversées constituent un palindrome.

Séquences répétées inversées terminales : courtes séquences apparentées ou identiques qui sont présentes en orientation opposée aux extrémités de certains transposons.

Sexe hétérogamétique : a la constitution chromosomique 2A + XY.

Sexe homogamétique : a la constitution chromosomique 2A + XX.

Site accepteur d'épissage : jonction entre l'extrémité droite d'un intron et l'extrémité gauche d'un exon.

Site actif : partie d'une protéine à laquelle se lie un substrat.

Site ATT : locus sur un phage et sur le chromosome bactérien au niveau duquel se produit la recombinaison qui permet l'intégration ou l'excision du phage à partir du chromosome bactérien.

Sites de remplacement dans un gène : sites où des mutations altèrent l'acide aminé qui est codé.

Site d'hypersensibilité à l'ADNase : courte région de chromatine détectée par son extrême sensibilité au clivage par l'ADNase I et d'autres nucléases ; il correspond probablement à une région d'où les nucléosomes sont exclus.

Site donneur d'épissage : voir jonction gauche d'épissage.

Sites silencieux dans un gène : décrivent les positions où des mutations n'altèrent pas le produit du gène.

Snurps : petites ribonucléoprotéines nucléaires (ARNsn associés à des protéines).

Sous-enroulement de l'ADN : produit par superenroulement négatif (car la double hélice est elle-même enroulée dans le sens opposé par rapport à l'entrecroisement des deux brins).

Southern blotting : décrit le protocole de transfert de l'ADN dénaturé d'un gel d'agarose sur un filtre de nitrocellulose où il pourra être hybridé à un acide nucléique complémentaire.

Sphéroplaste : cellule bactérienne ou de levure dont la paroi a été en partie ou totalement éliminée.

Sporulation : production d'une spore par une bactérie (par conversion morphologique) ou par une levure (comme produit de méiose).

Squelette d'un chromosome : structure sans protéine dans la configuration d'une paire de chromatides sœurs, engendrée lorsque les chromosomes sont dépourvus d'histones.

SSB : protéine d'*E. coli* se liant à l'ADN simple brin.

Streptolydignes : inhibent l'élongation de la transcription par l'ARN polymérase bactérienne.

Structure chi : jonction entre deux molécules double brin d'ADN, révélée en clivant un intermédiaire de deux cercles unis qui engendrent des extrémités linéaires dans chaque cercle ; elle ressemble à la lettre grecque chi d'où son nom.

Structure co-intégrée : structure produite par la fusion de deux réplicons, l'un possédant à l'origine un transposon, et l'autre en manquant. Les structures co-intégrées ont des copies du transposon présentes aux deux jonctions du réplicon, orientées comme des répétitions directes.

Structure quaternaire d'une protéine : fait référence à sa constitution multimérique.

Structure tertiaire d'une protéine : décrit l'organisation dans l'espace de sa chaîne polypeptidique.

Substitutions neutres dans une protéine : changements d'acides aminés qui n'affectent pas son activité.

Superenroulement : décrit l'enroulement d'un ADN double brin fermé dans l'espace de telle façon qu'il traverse son propre axe.

Superenroulement de l'ADN : il est provoqué par des supertours positifs (qui appliquent une tension plus grande dans la direction d'enroulement des deux brins l'un par rapport à l'autre dans le double brin).

Super-réprimé : signifie la même chose que non inductible.

Supertours négatifs : rendent compte de la torsion d'un double brin d'ADN dans l'espace, dans le sens opposé à celui des tours effectués par les brins d'ADN le long de la double hélice.

Supertour positif : décrit un enroulement de la double hélice dans l'espace dans la même direction que l'enroulement des deux brins de la double hélice.

Suppresseur ambre : gène muté qui code pour des ARNt dont les anticodons ont été modifiés de telle sorte qu'ils peuvent répondre tout aussi bien aux codons UAG qu'à leur précédent codon.

Suppresseur de non-sens : gène codant pour un ARNt mutant capable de se lier à l'un ou plusieurs des codons non-sens, entraînant à ce niveau l'addition d'un acide aminé supplémentaire.

Suppresseur d'ocre : gène codant pour un ARNt mutant capable de répondre au codon UAA pour permettre la poursuite de la synthèse protéique ; les supresseurs d'ocre suppriment également les codons ambres.

Suppresseur (extragénique) : est généralement un gène qui code pour un ARNt mutant qui lit le codon muté, soit dans le sens du codon initial, soit en donnant un substitut acceptable à la signification originale.

Suppression : décrit l'apparition de modifications qui éliminent les effets d'une mutation sans que soit supprimée la modification initiale dans l'ADN.

Synapse : décrit l'association de deux paires de chromatides sœurs correspondant à des chromosomes homologues, qui se produit au début de la méiose ; la structure en résultant est appelée un bivalent.

Synténique : les loci génétiques synténiques sont situés sur le même chromosome.

Synthèse d'ADN non programmée : toute synthèse d'ADN se produisant hors de la phase S du cycle cellulaire eucaryote.

Synthèse protéique cytoplasmique : traduction des ARNm qui représente les gènes nucléaires ; elle s'effectue via les ribosomes liés au cytosquelette.

Taux de mutation : taux avec lequel une mutation particulière apparaît ; il est généralement donné comme le nombre d'événements par gène et par génération.

Télomère : extrémité naturelle d'un chromosome.

Température de fusion (Tm : « melting temperature ») : point médian de la courbe de température au-dessus duquel l'ADN est dénaturé.

Termineur : séquence d'ADN située à la fin d'un transcrit, qui provoque la terminaison de la transcription par l'ARN polymérase.

Termineurs rho-indépendants : séquences d'ADN qui permettent à l'ARN polymérase de *E. coli* de terminer *in vitro* la transcription en l'absence du facteur rho.

Test cis-trans : détermine l'effet de la configuration relative sur l'expression de deux mutations. Chez un double hétérozygote, deux mutations dans le même gène conduisent à un phénotype muté dans la configuration *trans*, sauvage dans la configuration *cis*.

Test de croisement : croisement d'un génotype inconnu avec un homozygote récessif de telle sorte que les phénotypes de la descendance correspondent directement aux chromosomes portés par le parent de génotype inconnu.

Thalassémie : maladie des globules rouges du sang résultant de l'absence ou de la diminution, soit de l'alpha, soit de la bêta-globine.

Tige : segment apparié d'une épingle à cheveux.

Tm : abréviation pour température de fusion.

Topo-isomérase : enzyme qui peut modifier le « linking number » de l'ADN (par étapes de 1 pour celles de type I, par étapes de 2 pour celles de type II).

Traceur : composant d'acides nucléiques marqué radioactivement, qui est inclus dans une réaction de réassociation à des taux trop faibles pour influencer le déroulement de la réaction.

Traduction : synthèse de protéine sur une matrice d'ARN.

Traduction inverse : technique pour isoler les gènes (ou les ARNm) grâce à leur capacité à s'hybrider avec une courte séquence oligonucléotidique préparée en prédisant la séquence d'acide nucléique à partir de la séquence protéique connue.

Transcription : synthèse d'ARN sur une matrice d'ADN.

Transcription divergente : se réfère à l'initiation de la transcription à partir de deux promoteurs se faisant face, mais orientés en sens inverse, de telle sorte que la transcription s'effectue dans les deux directions à partir d'une région centrale.

Transcription inverse : synthèse d'ADN à partir d'une matrice d'ARN ; elle est réalisée par la transcriptase inverse.

Transcrit primaire : produit d'ARN initial non modifié correspondant à l'unité de transcription.

Transduction : fait référence au transfert d'un gène bactérien d'une bactérie à une autre par un phage ; un phage portant les gènes de l'hôte ainsi que les siens est appelé un phage transducteur. Elle décrit également l'acquisition et le transfert de séquences cellulaires eucaryotes par des rétrovirus.

Transfection de cellules eucaryotes : acquisition de nouveaux marqueurs génétiques par incorporation d'ADN ajouté.

Transformation de bactéries : décrit l'acquisition de nouveaux marqueurs génétiques par incorporation d'ADN ajouté.

Transformation de cellules eucaryotes : fait référence à leur conversion à un état de croissance en culture non restreint, ressemblant ou étant identique à la condition tumorigène.

Transition : mutation dans laquelle une pyrimidine est substituée à une autre ou dans laquelle une purine est substituée à une autre.

Translocation d'un chromosome : décrit un réarrangement dans lequel une partie d'un chromosome détachée par cassure s'attache à un autre chromosome.

Translocation d'un gène : fait référence à l'apparition d'une nouvelle copie en une localisation différente dans le génome de celle de la copie originale.

Translocation du ribosome : mouvement qu'effectue un codon le long de la chaîne d'ARNm après addition de chaque acide aminé à la chaîne polypeptidique en cours de synthèse.

Translocation réciproque : échange d'une partie d'un chromosome avec une partie d'un autre chromosome.

Transposase : activité enzymatique impliquée dans l'insertion d'un transposon à un nouveau site.

Transposon : séquence d'ADN capable de se répliquer et d'insérer une de ses copies à un nouveau site dans le génome.

Transposon composé : transposon ayant une région centrale bordée de chaque côté par des séquences insérées dont l'une ou les deux peuvent rendre l'élément entier capable de se transposer.

Transvection : décrit la capacité d'un locus à influencer l'activité d'un allèle sur l'autre homologue, uniquement lorsque deux chromosomes sont en synapse.

Transversion : mutation dans laquelle une purine est remplacée par une pyrimidine ou vice versa.

Unité de répétition dans un groupe en tandem : longueur de la séquence qui est répétée ; elle apparaît sous forme circulaire sur une carte de restriction.

Unité de transcription : distance entre les sites d'initiation et de terminaison de l'ARN polymérase ; elle peut comprendre plus d'un gène.

Unité formant foyer (uff) (en anglais ffu) : mesure quantitative de la formation de foyers cellulaires.

Valeur C : quantité totale d'ADN dans un génome haploïde.

Variation de phase : décrit une alternance dans le type de flagelles produits par une bactérie.

Vecteur de clonage : plasmide ou phage utilisé pour porter un ADN étranger inséré, dans le but de produire plus de matériel ou un produit protéique.

Vecteur d'expression : vecteur de clonage désigné ainsi car une séquence codante insérée en un site particulier sera transcrite et traduite en protéine.

Vecteur navette : plasmide construit de telle façon qu'il possède les origines de réplication, actives dans deux hôtes (par exemple, *E. coli* et *S. cerevisiae*), de telle sorte qu'il peut être utilisé pour véhiculer une séquence étrangère dans les procaryotes et les eucaryotes.

Virion : manteau externe protéique d'une particule virale.

Virus auxiliaire (helper) : apporte les fonctions absentes d'un virus défectif, permettant à ce dernier de compléter son cycle infectieux durant une infection mixte.

Virus défectueux dans la réplication : a perdu un ou plusieurs gènes essentiels pour achever son cycle infectieux.

Writhing number (nombre d'enroulements) : nombre de fois où l'axe d'un double brin se croise dans l'espace.

Zygote : est produit par la fusion de deux gamètes — c'est donc un œuf fertilisé.

INDEX

Acentrique :

- chromosome, 509
- fragment, 609

Acétylation :

- acides aminés, 7
- histone, cœur d'une, 547-548
- histones, modification des, 546-549
- lysine, 547
- nucléosomes, assemblage des, 547

Acides aminés, 4

- ARNt, chargement sur les, 98-100
- classification, 5-6
- discrimination (action de la synthétase), 129-130
- enchaînement dans les protéines, 4-8
- modification, 5-6
- oncogène-proto-oncogène, conversion, 704-705
- polaire, 5
- voir aussi* les acides aminés particuliers

Acide désoxyribonucléique, *voir* ADN

Acide gras, 4

Acide nalidixique, 582

Acide nucléique :

- association, 58-59
- bases azotées, 41
- classification, 4
- dogme central, 97-98
- hybridation, 58-60
- région codante, 87
- répétitions inversées, 63-65
- réplication double brin, 91

Acide nucléique (*suite*)

- simple brin :
 - réplication, 91
 - structure secondaire, 60-63
 - synthèse, 44-45
 - taille, anomalie de, 499-500
 - topologie, 57-71

Acide ribonucléique, *voir* ARN

Acridine, 51-52

Actine, structure des gènes d', 394-395

Activateur catabolique, protéine, *voir* protéine CAP se liant à la coiffe

Adénine, 41

Adénovirus :

- ADN, 332-334
- expression tardive, 390-391
- fonction, 700

Adénylate cyclase, 238

ADN :

- adénovirus, 332-334
- ARNm, copiant, 358-359
- ARN polymérase, 189-191
- empreinte (footprinting), 196-200
- Int, sites, 585
- modèle de la double hélice, 280-281

voir aussi Séquence consensus

biologie :

- flexibilité structurale, 555-557
- manipulation, 578-580
- nombre de torsion, 70
- boucle D, formation de la, 305-307
- brin L, 429-430

ADN (*suite*)

- brin H, 429-430
- carrier, 633
- chaîne, formation d'une, 579
- chimérique :
 - construction, 355-358
 - coupures décalées, 355-357
 - extrémités cohésives, 355-357
 - liaison à bouts francs, 357-358
 - poly (dA : dT), technique d'addition des queues, 357-358
- chi, structure, 571
- chloroplaste, 423-424
- chromosomique, 361-362
- cartographie génétique, 82-83
- circulaire, 68-70
- molécules, 423-424
- molécules ouvertes et circulaires, 505
- phage, 320-322, 582-583
- recombinaison réciproque, 569-570
- viral, 612-614
- clivage :
 - bandes, 75
 - digestion, technique de double, 75-78
 - enzymes de restriction, 339-340
 - fragment chevauchant, 76
 - gel, calibrage par, 75
 - hybridation des acides nucléiques, 78
 - marqueurs, 75

ADN (*suite*)

distribution des fragments, 79-80
 cœur, 524
 schéma de numérotation, 527
 complémentaire, voir ADNc
 composants, 41-44
 sites de coupure, 527
 concatémères, 501-502
 contenu, 411-412
 coupures, double brin, 526-527
 cruciforme, 63-65
 dans les gènes, 411-412
 découverte, 37-39
 dénaturation, 57-58
 détérioration :
 définition, 342
 E. coli, 345-346
 mutations ponctuelles, 342
 recombinaison-réparation, enzymes de, 344
 réparation post-réplivative, 344
 réponse SOS, 348-350
 structure, distorsions de, 342
 systèmes de recouvrement, 342
 digestion limitée, 524
 divergence des séquences, 406-408
 double brin :
 fils, 49
 hétérologue :
 boucle D, 568
 cassure et réunion, 565-568
 conversion génique, 577-578
 modèle de cassure double brin, 568-569
 homologue, 564-565
 parental, 49
 réplication, 325-329
 structure, flexibilité de la, 556
 structures en double hélice, 65-67
 double hélice, 45-48
 droite, 48
 grand sillon, 48
 petit sillon, 48
 séparation des brins, 578
 structures alternatives, 65-67
 super enroulement et structure, 70-71
 égoïste, 590
 étranger :
 définition, 336
 plasmide, insertion dans un, 358
 vecteur de clonage, 354
 excision réparation, système d', 342-344
 extrachromosomiques, copies, 632

ADN (*suite*)

famille Alu, 617-618
 forme A, 66
 forme B, 48, 66
 structures en double hélice, 65-67
 topologie, 57
 forme réplivative (RF), 320-322
 forme Z, 65-68
 gènes, dans les, 411-412
 gradient de densité, 433
 hélice de reconnaissance, 281
 hybride, 565
 information génétique, 90-93
 kinétoplaste, 557
 liaison :
 ARN polymérase, 189-191
 footprinting, 196-200
 coopérative, 556
 dissociation, 231-233
 modèle de la double hélice, 280-281
 région N-terminale, 230
 répresseur protéique, 227-231
 sites, Int, 585
 liaison à bouts francs, 357-358
 linker, 524-525
 marqueurs, 75
 marqueurs, des acides nucléiques,
 hybridation, 78
 enzymes de restriction, 73-75
 fragments chevauchants, 76
 profil des fragments, 79-80
 migration de branche, 565-568
 simple brin, 566, 568
 mitochondrial de levure, 430-431
 mitochondrial humain, 428-430
 molécules double-brin filles, 49, 539-540
 molécules double-brin hétérologue :
 boucle D, 568
 cassure et réunion, 565-568
 conversion génique, 577-578
 modèle de cassure double brin, 568-569
 molécule de liaison (linker), 524-525
 nodule (puffs), 626
 nœuds, 579
 non répétitif :
 gènes interrompus, 392
 gènes de structure, 377-378
 génom, taille et complexité du, 372
 renaturation, 375-377
 nucléosome, échelle, 522-524
 opérateur, 223-225
 organisation hiérarchisée, 519

ADN (*suite*)

palindrome, 63-65
 paradoxe de la valeur C, 367-368
 périodicité :
 coupure, 528-529
 structurale, 528-529
 superenroulement et, 530-531
 plasmide Ti-génom, interaction, 675-680
 proviral
 rare (complexe), 382
 réaction conduite pour l', 377
 réarrangement, trypanosome, 670-675
 réassociation, cinétique de, 371-372
 recombinant, 353
 récupération, système de, 336
 renaturation, 57-58
 répétitif : 372-373
 bande principale, 433
 hautement, 373
 satellite formation, 433-434
 modérément, 373-375, 432-433
 renaturation, 375-377
 répétitions inversées, 63-65
 réplication semiconservatrice, 49
 satellite, voir ADN satellite
 sauvegarde, 335-350
 échange de simple brin, 344
 restriction et modification, 336-338
 séquençage, 83-85
 ajustabilité, 625
 clonage bactérien, 354-355
 codant des protéines multiples, 89-90
 complexité, 369-371
 échelle, 84
 fidélité, 85
 fragments chevauchants, 85
 gènes interrompus, 87-88
 méthode chimique, 84-85
 mutations, 37
 nucléosomes, mise en phase des, 531-533
 voir aussi Consensus, séquences
 séquences, divergence des, 406-408
 évolution, horloge de l', 408-409
 sites de remplacement, 407
 sites silencieux, 407
 simples, organisation des séquences, 432-442
 sites régulateurs, 556-557
 sous-enroulé, 69
 structural, 528-529
 structure des familles, 65-66

ADN (*suite*)

- structure fermée, superenroulement, 68-70
- superenroulé *voir* ADN superenroulé
- super enroulement :
 - négatif, 505
 - octamère d'histone, 525-530
 - recombinaison spécifique de sites, 584
- voie non retenue, 505-506
- voie retenue, 505-506
- voir aussi* ADN, super enroulements
- supertours négatifs, 69-70, 505
- gyrase, 580-582
- sur torsadés, 69
- synthèse : 300-303
 - brin guide, 317
 - brin retardé, 317
 - mouvement du primosome, 324-325
 - brin viral, 320-322
 - contrôle de lecture, 315
 - discontinue, 316-317
 - initiation des simples brins, 320-322
 - non programmé, 350
 - nucléotides, 314
 - période G1, 300-301
 - période G2, 301
 - phage T4, 330
 - phase S, 301
 - primosome, mouvement du, 322-325
 - systèmes de réparation, 336
 - courts, fragments de, 345
 - erreur de, 347
 - excision-réparation, 342-345
 - longs fragments de, 345
 - mammifères des, 350
 - terminaison du brin codant, 41-44
 - topologie :
 - sensibilité structurale, 557-558
 - manipulation, 578-580
 - nombre de torsion, 70
 - transfected, 634-636
 - universalité de, 39-41

ADNase :

- coupure, 535
- purification, 39
- sensitivité chromosomique, 546
- sites hypersensibles, 551-555

ADNase I :

- analyse de haute résolution, 527-528
- coupure, 526
- digestion de la chromatine, 545
- échelles, 527

ADNase I (*suite*)

- particule cœur, 529-530
- sites hypersensibles, 552-554
- ADNase II :
 - coupure, 526
 - échelles, 527
 - sites hypersensibles, 552-554
- ADN-B, 557
- conversion en ADN-Z, 71
- ADNc, 359
 - carte de restriction, 385-386
 - chromosome polytène, 517
 - clonage, 359
 - excision-épissage, 473-476
 - hybridation, 361
 - isolement d'un gène, 359
- ADN de liaison, 524-525
- ADN ligase, 318-319
 - excision-réparation, système d', 343
 - fragments d'Okazaki, 312
- ADNm, 383
- ADN polymérase, 312
 - activités enzymatiques, 313-314
 - dimères de thymine, 344
 - propriétés, 313
- ADN polymérase α , 313
- ADN polymérase I, 313
 - amorce, 322
 - fragments d'Okazaki, 318
 - propriétés, 315-316
- ADN polymérase II, 313
- ADN polymérase III, 313
 - amorce ARN, 322
 - coupure, 329
 - sous-unités, 316
- ADNr, 415
 - extrachromosomique, 512-513
 - transcription, 416-417
- ADN satellite :
 - arthropode :
 - brin lourd (H pour heavy), 435
 - brin léger (L pour light), 435
 - répétitions identiques, 435
 - cryptique, 433-434
 - fixation du croisement, 442
 - hautement répétitif, 433-434
 - hétérochromatine, 434
 - mammifères, 435-438
 - souris, 436-438
 - évolution, étapes de l', 438-439

ADN super-enroulé :

- ADN fermé, 68
- bromure d'éthidium, 535
- chromatine, 530-531
- efficacité d'un promoteur, 200
- négative, double hélice, 69-71

ADN super-enroulé (*suite*)

- positive, 69
- site spécifique de recombinaison, 584
- structures de la double hélice, 70-71
- voies contraintes et non limitées, 505-506
- ADN-T, 676-677
 - nopaline, 678
 - processus de transfert, 679
 - répétitions directes, 678
- ADN topo-isomérase, 579-581
- ADN-Z, 65-68
 - ADN-B, conversion en, 71
 - enhancers, 215
- Agrobacterium tumefaciens*, 675-677, 680
- Agropine, plasmide, 676
- Allèle
 - codominant, 20
 - définition, 19
 - dominance incomplète, 20
 - dominant, 20
 - indépendance, 73
 - Mat*, 665-666
 - multiple, 33
 - récessif, 20
 - ségrégation, 19-20
 - type sauvage, 32
- Allélique, exclusion, 653-655
- Allolactose, 234
- Allostérique (contrôle allostérique, répresseur protéique), 223
- Alternance :
 - période courte, 376-377
 - période longue, 376-377
- Alu, famille, 617-618
- α -Amanitine, 188
- Ambre, codon, 141
- Amorce, ARNm-ADN copiant, 358-359
- AMPc :
 - répression des catabolites, 238-240
 - transcription, initiation de la, 241
- AMPcyclique, *voir* AMPc
- Amplification :
 - gène, 625
 - modèle, 626-627
 - oncogène, 706-711
 - séquences génomiques, 628-632
- Anasarque foeto-placentaire, 402
- Ancrage, dépendance de l', 698
- Anémie falciforme, liaison à un gène, 32
- Antennapedia*, groupe, 696-697
- Anti-association, facteurs d', 108
- Antibiotiques, synthèse protéique et, 152-153

Anticodon :

- anticodon-codon, reconnaissance, voir codon-anticodon, reconnaissance
- définition, 98-99
- mutants ARNt, 137-140
- oscillation (wobble), hypothèse de l', 132-133
- précision, 157
- suppresseur ambre, 138, 141
- suppresseur ocre, 120, 138-139, 141

Anticorps :

- hybridomes, 658
- mutation somatique, 658
- réponse immune, 641
- voir aussi Immunoglobuline

Anticorps-antigène, réaction, 645

Antigène :

- histocompatibilité, 642
- multiplication clonale, 644
- mutation somatique, 658
- réponse immune, 641
- transplantation, 660-661
 - domaine cytoplasmique, 661-662
 - domaine externe, 661-662
 - région transmembranaire, 662
- trypanosome, réarrangement de l'ADN, 670-675
- voir aussi Antigène d'histocompatibilité

Antigène d'histocompatibilité, 642

- loci, 660-663
- $\beta 2$ microglobuline, 662-663

Antigénique, variation, 670-675

Antirépresseur, infection lytique, 287-289

Antiterminaison :

- cascade du phage lambda, 274-276
- définition, 249
- facteurs, 249
- protéine, 276
- spécificité du site, 261-266

Aplysie, 495

Apo-inductrice, protéine, 236

Appariement, voir Appariement des bases, Synapses

ARN :

- activité catalytique, 461-465
- amorce :
 - fragments d'Okazaki, 317-318
 - phage T4, 330
- (dot blotting), transfert en tache, 361-362
- double hélice, 61-63
 - alternatives structurales, 66-67

ARN (suite)

- boucle, intérieur d'une, 63
- boucles en épingle à cheveux, 62-63
- boucles en protubérance, 63
- double-brin, 244-245
- épissage, 458-482
 - activité catalytique, 461-465
 - auto-épissage, 462-463
 - complexe du spliceosome, 477-478
 - éléments Ty, 619-620
 - gènes interrompus, 88, 386-387
 - jonctions :
 - interchangeabilité, 473-475
 - intermédiaires en lasso, 475-476
 - pré-épissage, complexe de, 477-478
 - réactions apparentées, 479-482
 - spécificité tissulaire, 474
 - systèmes, 458
 - systèmes *in vitro*, 474-478
- expression des chaînes lourdes, 656-657
- gènes, expression des, 183-184
- information génétique, 90-93
- Northern blotting, 361
- point de terminaison, 248
- poliovirus, 332-334
- PSTV, 91-93
- p16 et p22, 452-453
- régulation de la traduction, 244-245
- rétrovirus, 612-613
- rho, facteur, 260-261
- structure de, 41
- structure primaire, 60-61
- superflus, accessoires, 446-448
- synthèse, réplication, 308-310
- TMV, trajet en hélice, 501
- traceur, 377-378
- viroïdes, 91-93
- ARN-ARN, réactions, 481
- ARNase D, 455
- ARNase H, 308-309
- ARNase P, 455
- ARNase I, 308-310
- ARNase II, 455
- ARNase III :
 - ARNr, maturation des, 452, 453
 - phage T7, ARNm précoce du, 448-450
- ARN ligase, 459-460
- ARNm :
 - abondance, 381-382
 - ADNc, polyadénylation des, 168-169
 - anti-sens, 245-247

ARNm (suite)

- appariement des bases :
 - ARNr, 171-172
 - Shine Delgarno Séquence, 171
- ARNt, groupe de gènes, 454-455
- bactérien, 101-103
 - instabilité, 160
 - polycistronique, 161-163
 - régions codantes, 162
- brin codant, 97
- brin non-codant, 97
- cartographie de restriction, 385-387
- c-myc*, 708
- complexe, 382
- copiant l'ADN, 358-359
- définition, 86-87
- demi-vie chimique, 160
- épissage, 469-470
- eucaryotique :
 - coiffe méthylée, 169-171
 - définition fonctionnelle, 164-166
 - migration des sous-unités, 172-173
 - polyadénylé, 168-169
 - vs. procaryotique, 159
- expression des gènes des chaînes lourdes, 656-657
- gène d'amylase de souris, 389-391
- généralités, 159
- gènes interrompus, 87-88
 - organisation, 384-385
- gènes de structure, 377-379
- hnRNA, dérivation, 485-487
 - comparaison des tailles, 484-485
 - contrôle post-transcriptionnel, 489-491
- hybridation :
 - ARNm $\times$ ADNc, 380-381
 - non répétitive, 379
 - saturation, chevauchement, 382-383
 - traceur, 378
- lac*, 223
- liaison ribosome-complexe d'initiation, 106, 108-109
- demi-vie fonctionnelle, 160
- monocistronique, 162
- objectif, 159
- ompF*, 245-247
- phage T7, 448-450
- polarité, 164
- polycistronique, 161-163
 - traduction, 163-165, 243
 - polarité, 164
- populations chevauchantes, 382-383
- rare, 382

ARNm (*suite*)

- région codante, 162
- "leader", 162
- translation, 163-165
- "trailer", 162
- régions intercistroniques, 162-163
- représentation, 381-382
- ribosome, fidélité du, 156-157
- susceptibilité à l'ADNase I, 546
- traduction, 160-161, 223
 - contrôle, 166-167
 - gènes ARNt, groupes de, 454-455
 - systèmes acellulaires, 166-167
 - systèmes *in vitro*, 166-167
- traduction cytoplasmique, 492
- traduction ribosomale, 99-103
- transcription, 160-161
- transitoire, 159-161
- viral, 173

ARN maturase, 469-471

ARN messagers, voir ARNm

ARN nucléaires, petits, voir SnARN

ARN polymérase :

ADN, liaison à l', 189-191

- comme non histone, 520
- complexes binaires, 190
- complexes fermés, 189-190
- complexes ternaires, 190-191
- empreinte (footprinting), 196-200
- facteur sigma, 189-191
- site de liaison perdu, 189-190

sous-unités eucaryotiques, 188-189

 α -amanitine, 188

antiterminaison, 261-262, 265

bactérienne :

- enzyme cœur, 185-188
- facteur sigma, 185-188
- sous-unités, composition des, 185-186

cœur, topologie de l'enzyme, 190-191

contact avec le promoteur, 192-199

séquences consensus, 193-195

contrôle atténuateur, 253-257

facteur rho, 261

interactions avec le promoteur,

initiation, contrôle, 183-218

liaison :

ADN, sites, 189-191

promoteur, 195-200

site de liaison CAP, 239-240

sous-unités :

antiterminaison, 267-268

sous-unités β , 239-240

organisation de l'opéron, 242-243

ARN polymérase (*suite*)

- régulation autogène, 244
- sporulation, 202, 204-205
- SP6, 364
- synthèse, séquence en épingle à cheveux, 320, 322
- terminateur, 248-251
- transcription :
 - catalyse, 184-185
 - phage lambda, 265
 - régulation positive, 286-287

ARN polymérase I, 188

terminaison, 251

ARN polymérase II, 188

chromosome puffing, 517

promoteur, 206-209

composants de la séquence, 210-211

facteurs de transcription, 211-213

structure multiple, 209-211

taille, anomalie de, 210

terminaison, 251

ARN polymérase III, 188

gènes 5S, 418-419

promoteur, aval, 216-218

terminaison, 251

ARNr :

appariement des bases, ARNm, 171-172

clivage chez les procaryotes et les eucaryotes, 450-454

définitions, 100

fonctions, 144-145

gènes répétés, 415-416

introns dans les gènes codants, 388

longueur, variations de, 148

mitochondriale, 148

opérons *rm*, 451-452

protéines ribosomales, 147-150

structure, 148-150

synthèse :

organisation, 418

réponse forte, 240-241

ARN ribosomal, voir ARNr

ARNr 16S :

gènes répétés, 415

sous-unités, 154-156

structure, 148-150

synthèse protéique, 153

ARNr 23 S :

gènes répétés, 415

peptidyl transférase, activité, 156

sous-unités, 154-156

structure, 149, 150

ARNr 35 S, épissage, 461-462

ARNt :

- adaptateur transcriptionnel, 122-142
- anticodon, 98-100
- apparenté, 126
- bras, 98-99
- classe 1, 123
- classe 2, 123
- correspondant, 126
- désacylés, 106
- discrimination à l'étape de charge, 128-129
- enzyme de maturation, 455-457
- excision-épissage :
 - levure, 459-461
 - in vitro*, 459-460
 - mammifères, 461
- terminaisons, 5' et 3', 460-461
- gènes, groupes de, 454-457
- gènes répétés, 415-416
- initiateur, 109
- introns dans les gènes codant, 388
- isoaccepteur, 122
- mitochondriaux, ensemble des, 137
- modification du codon, 135-136
- mutant, codon de lecture, 137-140
- promoteur, 216-217
- reconnaissance, anticodon, 129, 131-133
- reconnaissance par les synthétases, 126-128, 131
- rm*, 451-452
- séquences intercalaires, 452
- structure en feuille de trèfle, 98-100, 123-124
 - boucles, 98-99
 - bras, 123-124
 - accepteur, 123-124
 - anticodon, 123-124
 - boucles, 123
 - branches, 123
 - bras supplémentaire, 123-124
 - T ψ C, 123-124
 - liaisons H secondaires, 125
 - semi-invariant (semi-conservé), 123
- structure secondaire, voir ARN,
 - structure en feuille de trèfle
- structure tertiaire, 123, 125-126
- liaison H tertiaire, 125
- modèle de l'espace à remplir, 125-126
- suppresseur :
 - ambre, 138-139
 - codons, compétition des, 140-141
 - lecture au-delà, 141
 - léthalité récessive, 138-139

- ARNt (*suite*)
 ocre, 138-139, 141
 suppression d'un décalage du cadre, 142
 synthèse protéique, 98-99
 synthèse, réponse forte, 240-241
voir aussi ARNt spécifiques
- ARNt-aminoacyl, 106
 atténuation, 252
 puromycine, 119
 suppression, 142
 synthèse protéique :
 élongation, 114-116
 facteurs, 114-115
 initiation, 110
- ARNt-aminoacyl synthétase, 98-100, 126-128
 réponse stringente, 240-241
- ARNt Asp, 125
- ARNt f Met, 109-112
 inhibition par la kasugamycine, 153
 reconnaissance codon-anticodon, 135
- ARNt His, 258-259
- ARNt Ile, 129, 131
- ARNt isoaccepteur, 122
- ARNt Lys, suppression d'un décalage du cadre de lecture, 142
- ARNt nucléotidyl transférase, 457
- ARNt Phe, 125
 modification des bases, 133
 suppression d'un décalage du cadre de lecture, 142
- ARN de transfert, *voir* ARNt
- ARNt Ser, 139
- ARNt Trp, 139
- ARNt₂ Glu, 452
- ARN-Val, 109
- ARN 5S :
 gènes répétés, 415
 séquence promoteur, 216-218
 sous-unités 50S, 156
 structure, 148-150
- ARN 5,8S, structure, 149-150
- ARN 7S, particule de reconnaissance du signal, 175-177
- Arrêt du ribosome, 253
- Ars, séquences, 303-304
- Ascomycètes, recombinaison, 576-577
- Aspergillus nidulans*, génome mitochondrial, 425-427
- Assemblage des protéines, 503, 525
 assemblage des nucléosomes, 541
- ATP :
 ARNt, épissage in vitro, 459-460
 hydrolyse :
- ATP (*suite*)
 ADN super-enroulé, 580
 renouvellement enzymatique, 582
 sites cibles sur l'ADN, 339-340
 ligation, réaction de, 459-460
 simple-brins, assimilation des, 323-324
 topo-isomérisation, réaction de, 579-581
- ATPase
 mitochondrie de levure, 426-427
 primosome, mouvement du, 324
- Attachement (*att*), sites, 582
- Atténuateur, 253-254
- Atténuation (contrôle de l'atténuation) :
 arrêt du ribosome, 253-254
 définition, 253
 facteur rho, *E. coli*, 259-261
 généralité, 256-259
 mécanisme de l'histidine, 258-259
 mutation, 255-256
 action en *cis*, 258
 opéron tryptophane, 258-259
 séquences du peptide leader, 253-255
 mécanismes des acides aminés, 258-259
 structures secondaires, 252-256
 synthèse protéique, 256
- Autocatalyse, 463-463
- Auto-épissage, 462-463
 introns, 465-468
 réactions apparentées, 481
- Autogène, circuit :
 cascade, 287-288
 lysogénie, 276-278
- Autogène, contrôle, 234-235
 ARN polymérase, sous-unités de l', 244
 traduction, 242-244
- Autonomes, éléments, 606
- Autosomes, 26
- Bacillus subtilis*, facteur sigma, 201-206
- Backcross (technique du back cross ou croisement en retour), 22
- Background (niveau du background, mutations spontanées), 52
- Bactérie :
 ARNm, 101-103
 avirulente, 38
 clonage d'ADN, 354-355
 facteurs d'initiation, 11
 flexibilité, 219
- Bactérie (*suite*)
 hypothèse du signal, 176
 lisse, 38-39
 lyse, 11
 principe transformant, 38
 rugueuse, 38-39
 traduction, 159-161
 transcription, 159-161
voir aussi Eucaryotes
- Bactériophages, *voir* Phage
- Balbani, anneaux de, 517-518
- Bam, îlots, 418
- Bandes G, technique de marquage des, 508-509
- Base, appariement des :
 acides nucléiques, hybridation des, 58-60
- ADN :
 super-enroulement, 70-71
- ARN, excision - épissage de l', 472-473
- ARN I, 309-310
- ARNm-ADN, 358-359
- ARNm-ARNr, 171-172
- ARNt :
 structure en feuille de trèfle, 123-124
 structure tertiaire, 125-126
 carte de restriction, 74
 complémentarité, 45-47
 réplication, 90-93
 transcription et catalyse par l'ARN polymérase, 184-185
 énergie libre, 61-62
 matrice, 49-50
 mutation ponctuelle, 52-53
 paires de kilobases, 74
 plan, 47
 règles, 61
 réplication, 90-93, 314
 snARN, 478-479
 squelette, 45-47
 tautomérique, 47
 transposons IS, 595-597
 trp, atténuation du, 256-258
- Base azotée :
 définition, 41
 hydrogène, liaison, 45-47
- Base, empilement, 46
- Base, mauvais appariement des, 52-54
- Base, modification des, 55-56
 ARNt, 133-134
 ARNt, codon, contrôle du, 135-136
- Bithorax, 690-693
- Bivalent, 23
- Bloom, syndrome de, 350

- Boîte GC, 211
transcription, facteur de, 212
« Boîte Pribnow », 194
- Boîte TACTAAC, 476
- Boîte TATA (Hogness), 208-209
- Borrelia hermsii*, variation antigénique, 675
- Boucle D, formation d'une, 305-307
compréhension de l'assimilation des
simples, 572-573
déplacement de, 332-334
protéine se liant aux simples, 320-322
retardé, synthèse de l'ADN, 317
séparation des, initiation de la
transcription, 200
- Bromo uracile (BUdr), 53-54
- CAAT (boîte) :
facteur CTF, 212
promoteur, 211
- Cadre de Lecture, 51
ADN mitochondrial humain, 429-430
éléments Ty, 620
réactions de jonctions, 649
- Capside :
assemblage, 501-503
empaquetage, 502
filamenteux, 500
génomés (acides nucléiques), 500-503
nucléation, centre de, 501
symétrie icosaédrique, 500-501
tête vide, 502-503
- Carte :
distance, 28
unités, 28-29
voir aussi, Cartographie génétique ;
Cartographie de restriction
- Carte des territoires présomptifs, 685
- Caryotype, 23
- Cascade transcriptionnelle, 202-203
- Cassettes, modèles des, type sexuel
chez la levure, 667-670
- Catalyse biochimique, 463-464
- CAT, dosage de l'activité, 365-366
- C-banding, technique, 510-512
- C. elegans*, 368
localisation des gènes, 72
- Cellule hôte, 18
- Cellule :
cycle, 22-26
réplication, 299-300
établie, 698
foyer, 703-704
- Cellule (suite)
interphase, 23
non tumorigène, 698
oncogènes rétroviraux, 701-703
primaire, 698
somatique, 23
transposition ARN-dépendante, 616-617
voir aussi Méiose ; Mitose
- Cellules T, récepteur des, 642
relation avec les immunoglobulines,
659-660
- Cellules végétatives, facteur sigma, 202
- Centromère, 509
chromatine de levure, structure de la,
555
définition, 23
kinétochore, 510
- Cercle roulant, 303-305
- Chaîne polypeptidique, 5
extrémité amino- ou N-terminale, 5
extrémité carboxyl- ou C-terminale, 5
hélice alpha, 9-10
hémoglobine, 32
liaison hydrogène :
hélice α , 9
structure en feuillets β , 9-10
ossature, 5-8
structure primaire, 7
synthèse protéique, 99-101
- Chiasma :
cassure et réunion, 564
chromosomes en écouvillon, 514
liaison, 27-29
localisation des gènes, 29
recombinaison, 29
- Chi, séquence, ADN *E. coli*, 353
- Chi, structure, 571
site spécifique de recombinaison,
584-585
transposition, 601
- Chlamidomonas reinhardtii*, 386
hérédité maternelle, 423
- Chloramphénicol, synthèse protéique,
153-154
- Chloroplaste :
ARN-polymérase, 189
génomés, 423-424
- Chromatides sœurs :
bivalent, 23
chromosomes en écouvillon, 514
crossing over inégal, modèle du,
recombinaison, 652-653
définition, 23
métaphase, 506-507
squelette, 537
- Chromatides sœurs (suite)
synapse ou appariement, 23
- Chromatine :
active, 538-558
ADN, profil des bandes d', 532
boucles, 535-536
composants protéiques, 520-521
définition, 22
domaines, 535-536
empaquetage, 499
interphase, 507-509
matrice nucléaire, 485
nucléosome, 521-524
mise en phase, 531-533
protéines non histone, 540
reproduction, 539-540
squelette, 535-597
structure, 519-537
activation des gènes, 557-558
sites hypersensibles, 551-558
structure en filament, 542
superenroulement stable, 542
trajet des fibres, 533-535
fibres de 10 nm, 533-534
fibres de 33 nm, 533-535
transcription, domaines sensibles à
l'ADNase, 545-546
- Chromocentre, 508
- Chromomère, 513
- Chromosome :
amplification, 513, 630-632
appariement, 564
bactérien, 296
bande C, 510-512
bandes, 515-517
hybridation *in situ* ou cytologique,
516-517
bandes G, technique des, 508-509
carte cytologique, 516-517
cassure :
élément Δs , 607-609
fragment acentrique, 607-607
cassure - fusion - pont, cycle, 609
centromère, 509
chiasma, 27-29
complexe synaptonémal, 564-565
crossing-over pendant la méiose, 27-29
dicentrique, 609
double minute, 629, 631
écouvillon, 513-515
empaquetage, 499
ensemble diploïde, 23
2n membres, 23
4n, 23
eucaryotique, 296, 509-512

Chromosome (*suite*)

- réplicon, 300-303
 - fourches multiples, 300
 - fragment acentrique, 509
 - géant, 515-517
 - gènes sur le, 26-27
 - génomés et, 499-518
 - haploïde, 23
 - hérédité et, 22-26
 - histones éliminés, 536
 - homologues, 23
 - hybridation, 363-364
 - interbandes, 516-517
 - kinétochore, 510
 - liaison, 27-28
 - marche sur le, 361-364
 - microtubules, 509
 - minichromosome, 530-531
 - mitose, 23-24, 507-509
 - multifourche, 300
 - Philadelphia*, 709
 - polytène, 513-517
 - état boursoufflé (puff), 517-518
 - réarrangement, productif et non productif, 653-655
 - recombinant non réciproque, 402
 - réplication, cycle de, 299-300
 - réplication, semi-conservatrice, 508-509
 - ségrégation, unité de, 295
 - sexuels, 26
 - sous-unité nucléosomale, 521-524
 - structure, arrêt de la transcription, 517-518
 - télomère, 511-512
 - transcription, 517-518
 - translocation, 707
- Cinétique de correction d'épreuve, 129
- Cis*, définition de la dominance en, 225-227
- Cis-trans*, test de complémentation, 33-34, 389-391
- Cistron :
- ARNm, 163-165
 - définition, 33-35
 - voir aussi Gènes
- Classe, commutation de, 655-656
- Régions S, 655
- Clonage :
- ADN bactérien, 354-355
 - ADNc, 359
 - ADN chromosomique ou génomique, 361-364
 - carte de restriction, 386
 - cosmide, 353
 - dot blotting, 361-362

Clonage (*suite*)

- pBR322, 354-355
 - phages, 355
 - vecteur de, 354
 - vecteur d'expression, 364-366
- Clonale, analyse :
- éléments de contrôle, 607-608
 - secteurs jumeaux, 610
- Clonale, théorie de la sélection, 644
- Clone, définition, 353
- Coconversion, 577
- Codon :
- anticodon, reconnaissance, 157
 - ARNt, 129, 131-133
 - ARNt, mutants, 138-140
 - ARNt mitochondriaux, ensemble des, 137
- compétition :
- ARNt suppresseurs, 140-141
 - suppresseurs ambre, 141
 - suppresseurs ocre, 141
- contexte, 140-141
- contrôle
- ARNt, modification des bases de l', 135-136
- définition, 49
- familles, 129
- initiation, 109
- contexte, 109
- non chevauchant, 49
- ocre, 120, 141
- opal, synthèse protéique, 120
- oscillation, hypothèse de l', 132-133
- paires, 123
- point de départ fixé, 51
- suppression du décalage du cadre de lecture, 142
- synonymes, 104
- synthèse protéique, 103-105
- terminaison :
- ocre :
 - mutants ARNt, 138-139
 - synthèse protéique, 120
 - vecteur de clonage eucaryotique, 364-366
- Codon—anticodon, reconnaissance :
- traduction, fidélité de la, 157
- Coévolution, 420
- Code génétique, 37
- ADN ou ARN, approvisionnement, 90-93
 - acides nucléiques, 92
 - ARN et ADN, approvisionnement en, 90-93
 - dégénérescence, 104
 - dogme central, 97-98

Code génétique (*suite*)

- information, 3-4
 - mitochondrie, 135, 137
 - rétrovirus, 700-701
 - sens, 103-105
 - triplets, 49-52
 - universalité, 105
- Coiffe (CAP), protéine se liant à la :
- ARNm eucaryotes, 172-173
 - facteurs dans la synthèse protéique, 112
 - répression des catabolites, 239-240
- Coiffe (mise en place de la coiffe - capping) :
- hnARN, 486-487
 - structure des ARNm eucaryotes, 169-170
 - transcrits primaires, 492
- Compatibilité, groupe de, 307
- Complémentation :
- interallélique, 35, 229-230
 - mutation MAT, 667-668
 - négative, 229-230
 - test de, 33
 - test de, *in vitro*, 320
- Complexe bithorax, 690-693
- Complexe de préinitiation, 217
- Complexité chimique, 372
- Complexité cinétique, 372
- hybridation, 380-381
- Concatémériques, molécules, 501-502
- Condensation, rapport de :
- ADN, organisation du nucléosome, 542-543
 - chromatine, structure de la, 519
 - définition, 499-500
 - euchromatine, 507-508
- Condensation, réaction de, et liaison peptidique, 4-5
- Conformation, définition, 11
- Consensus, séquence :
- ADN chimérique, 356
 - ADN satellite, 436-438
 - CAP, 239-240
 - crossing-over inégal, 440-442
 - définition, 194
- EcoB, 339
- énergie libre, 61-62
- enhancer :
- glucocorticoïdes, régulé par les, 215-216
 - immunoglobulines, 214
- épissage :
- jonctions, TACTAAC, boîte, 476
 - 473-475
 - séquence guide interne, 466

Consensus (suite)

- exon-intron, jonctions, 388-389
 - sites donneurs et sites accepteurs, 389
 - gène, reconstruction d'un, 660
 - gènes interrompus, profil d'épissage des, 390-391
 - hnARN, polyadénylation des, 487-488
 - immunoglobines, gènes des, 650-653
 - introns, mutations dans les, 468
 - promoteur, 193-195
 - Bacillus subtilis*, facteur sigma des, 205-206
 - boîte Pribnow, 194
 - boîte TATA (Hogness), 208-209
 - conservé, 194
 - TATAAT, 194
 - 10, 194, 198-200
 - 35, 194, 198, 200
 - recombinaison, site spécifique de, 584-585
 - réplication saltatoire, 438-439
 - répresseur, synthèse du, 286
 - Tn10, transposon, 594-596
 - transcription, facteurs de, 211-213
- Contrôles positifs, antiterminaison, 264-265
- Contrôle d'épreuve chimique, 129
- Conversion génique, 420
- coconversion, 577
 - recombinaison interallélique, 576-578
- Copia*, famille, 620-622
- Copies extrachromosomiques, amplification des gènes, 630
- Copies, nombre de :
 - copie unique, 307
 - incompatibilité entre plasmides, 307-311
 - multicopies, 307
- Cordycépine, 307
- Corépresseur :
 - définition, 220
 - opéron, contrôle de l', 236
- Correction d'épreuves :
 - chimique, 129
 - cinétique, 129
 - conformationnelle, 129
 - étape de chargement des ARNt, 129
 - synthèse d'ADN, 315
- Cosmide, 355
- Cot, courbe du, 370
- valeur du :
 - ADN, renaturation, 376

Cot (suite)

- réassociation, cinétique de, 370-371
- Cotransfection, 633
- Coupure :
 - ADNase, 535
 - ADNase I et II, 526
 - ADN cœur, sites, 527
 - ADN polymérase I, 315-316
 - ADN, topologie de l', 578
 - chromatine superenroulée, 542
 - empreinte (footprinting), 196-200
 - intermédiaires enzyme-AMP, 318-319
 - migration de la fourche, 566-567
 - nucléotide, 505-507
 - synthèse du brin principal "leading", 327-328
- Croissance, points de, voir Réplication, fourche de
- Crossing over :
 - méiotique, 27-29
 - voir aussi Crossing-over inégal
- Crossing-over inégal :
 - ADN satellite, 440-442
 - gènes, groupe de, 402-403, 421
 - liaison, 652-653
 - thalassémie, 402-405
- Cross over :
 - fixation du, 420-421
 - répétition identique, 442
- Cruciforme :
 - définition, 63-65
 - in vitro, 63-65
 - in vivo, 63-65
- Cyclisation :
 - primaire, 463
 - secondaire, 463
- Cytochrome b. épissage, 469
- Cytoplasme dans les cellules, 3
- Cytosine, 41
- désamination, 55
- Cytosquelette, définition, 173-174
- D. discoideum*, 368-369
- gènes interrompus, 388
 - molécules extrachromosomiques, 512-513
- Délétion :
 - description, 35-36
 - locus *white*, 682-683
 - mutations ponctuelles, 53
 - promoteur, liaisons au, 207
 - recombinaison de gène C et V, 650-653

Délétion (suite)

- sites de, 78-79
 - transposons, séquences de, 598
- Déméthylation, expression des gènes, 549-551
- Dénaturation :
 - ADN, 57-58
 - effet hyperchrome, 58
 - protéine, 11
 - voir aussi Renaturation
- Densité de saturation, 698
- Densité de superhélicité, 69
- Désoxyribose, 42
- Différenciation terminale, 23
- Digestion limite, 524
- Digestion de restriction, fragmentation du génome, 360
- Digestion, technique de, 75-78
- Dihydrofolate réductase, 629
- Dissociation, facteur de, 108
- D. melanogaster* :
 - ADN satellite, 435
 - ADN Z, 68
 - ARNm anti-sens, 245-247
 - banque génomique, 362
 - bithorax complexe, 690-695
 - BX-C mutation, 688-690
 - cellules transgéniques, 636
 - chromatine, fibres de, pendant la transcription, 544
 - chromatine, squelette de la, 537
 - chromosomes, centromères des, 516
 - facteurs P, 622-623
 - facteurs de transcription, 212
 - gène, cartographie d'un, 72
 - gènes histones, organisation des, 414-415
 - gènes interrompus, 388
 - gènes ordinaires, 693-694
 - génom, isolement du, 535-536
 - homeo box, 695-697
 - liaison, groupe de, 30
 - loci complexes, 689-695
 - loci, mutation des, 369
 - locus *white*, 682-685
 - méthylation, 551
 - polymorphisme génétique, 78-79
 - réplicons, 303
 - transposons, 620-622
 - UH2A, 548
 - unité, compartiment, 685
 - unité segmentale, 685
- Domaine chromosomique, sensibilité
- ADNase, 546
- Dot blotting, 361-362

- Dysgénésie hybride, 622-624
- Échelle :
- micrococcale, 544
 - nucléosome, 522-524
- EcoB, enzyme, 338-339
- EcoK, enzyme, 338-339
- E. coli* :
- ADN, analyse de l', *in vivo*, 506
 - ADN, protéines se liant à l', 504-505
 - ADN, synthèse de l', 319-320
 - ARN polymérase, sous-unités, 185-186
 - ARNr 16S, 148-149
 - colinéarité gène-protéine, 86
 - endonucléase RecBC, 575-576
 - épissage, réaction d', 467-468
 - facteur rho, 259-261
 - facteur sigma, 200-206
 - gènes ARNt, groupes de, 454-455
 - lysé-nucléotide, 504
 - opéron, 224-225
 - protéines ribosomales, 242
 - recombinaison-réparation, système de, 347-348
 - réparation-excision, système de, 345-347
 - ribonucléases, 447
 - structures ribosomales, 145-147
 - suppresseurs, 140
 - synthèse protéique, composants, 101, 103
 - termination, modes de, 249-250
 - tryptophane, 252
 - unités transcriptionnelles, 160-161
- EcoP1, 341
- EcoP5, 341
- EcoR1 :
- ADN chimérique, 356-357
 - restriction, 337-338
 - restriction, carte de, 78
- EcoR11, ADN satellite, 439-440
- EDTA, traduction des ARNm, 164-166
- EGF, voir Facteur de croissance épidermique
- Électrophorèse sur gel :
- ADN :
 - clivage, 74-75
 - fragments, 360
 - séquençage, 84-85
 - enzyme de restriction, 550
 - épissage des ARNt de levure, 459-460
 - nucléase micrococcale, 523
- Élément P, transfection, 636
- Éléments de contrôle, 605-607
- Ac éléments, 608-609
 - Ds élément, 607-609
- Éléments de contrôle du maïs, 605-607
- Éléments non autonomes, 606-607, 609
- Éléments transposables, voir Transposon
- ELH, neuropeptide, 495-496
- Elongation, facteur :
- eEF1, 116
 - eEF2, 116, 119
 - EF-G, 117-119
 - nécessitant une translocation, 117-119
 - protéines, synthèse des, 105
 - sites nucléotidiques, 186
- T, 114-116
- T, EF-Ts, 116
- EF-Tu, 116
- transcription :
- catalyse par l'ARN polymérase, 185
 - nucléotides, 241
- Énergie libre, 61-62
- Enhancer :
- bidirectionnel, 213-216
 - cellulaire, 214
 - définition, 214
 - liaisons, 214
 - rétrovirus, 214
- Enzyme :
- complète, voir Holoenzyme
 - définition, 4
 - HpaI, 549-550
 - modification, 335-336
 - recombinaison-réparation, enzymes de, 344
 - renouvellement, 582
 - restriction, voir Enzyme de restriction
 - transcription, inhibiteurs de la, 189
 - type I, topologie de l'ADN, 579
 - type II, topologie de l'ADN, 579
- Enzyme de restriction, 335
- ADN satellite, 439-440
 - clivage de l'ADN, 73-75
 - coupures en chicane, 355-357
 - EcoRI, 78
 - électrophorèse sur gel, 73-75, 550
 - extrémités collantes, 355-357
 - HpaI, 78
 - MspI, 549-550
 - nomenclature, 78
 - opération, 336-338
 - type I, 338-341
 - type I, sites cibles, 339-340
- Enzyme de restriction (*suite*)
- type II, 337-338
 - type III, 341-342
 - sous-unités, 341
- Épingle à cheveux :
- appariement ARNm-ARNr, 171-172
 - boucle, 60
 - bras, 60
 - définition, 60
- Épisomes, 270
- Épreuve conformationnelle, correction d', 129
- Espacement de période courte, 376-377
- Essai de liaison au filtre, 227
- Eucaryotes :
- ADN chez les, 39-41
 - ARNm, 159
 - coiffe méthylée, 169-171
 - définition de la, 164-166
 - migration des sous-unités, 172-173
 - polyadénylé, 168-169
 - ARN polymérase, 188-189
 - définition, 3
 - éléments mobiles, 605-624
 - expression des gènes dans les systèmes procaryotes, 364-366
 - gènes interrompus, 87-89
 - génomés, 367
 - recombinaison réciproque, 589
 - régulation autogène, 244
 - réplication, 313
 - synthèse protéique, initiation de la, 112-114
 - termination, 251-252
- Euchromatine, 434
- changements cycliques, 538
 - empaquetage, rapport d', 507-508
 - gènes actifs, 508
- Évolution concertée, 420
- Évolution, horloge de l', 408-409
- Excision :
- ADN lambda, 582
 - transposition, 598
 - imprécise, 598
 - précise, 598
- Excision/épissage, 458-482
- ARNt de levure, 459-461
 - box gènes, 469
 - en *trans*, épissage, 481-482
 - in vivo*, 482
 - gènes des chaînes lourdes, expression des, 656-657
 - jonctions :
 - interchangeabilité, 473-475

- Excision/épissage (*suite*)
intermédiaire, en lasso, 475-476
réactions apparentées, 479-482
in vitro, 480
snARN, 478-482
voir aussi Auto-épissage
- Excision-réparation, système d', 342-345
E. coli, 345-347
excision, 342
incision, 342
réparation, 342
- Exon :
complémentation intronique, 389-391
définition, 88
éléments P, 623-624
gènes interrompus, 385-387
évolution, 392
mutations, cartographies des, 682-685
et protéine, 395-396
- Exon-intron, jonction, 388-389
- Exonucléase, 198
action au hasard, 3445
action processive, 445
- Expérience de "Playback", 379
- Expression, copie liée à l', (ELC), 672-673
- Facteur de croissance dérivé des plaquettes (PDGF), 714
- Facteur de croissance épidermique (EGF), 694-695
récepteurs, oncogènes, 714
- Facteur particulier, 18-19
- Facteur de recyclage, synthèse protéique, 111
- Facteur sigma :
ARN polymérase bactérienne, 185-188
Bacillus subtilis, 201-206
promoteur, séquences consensus du, 205-206
E. coli, 200-206
liaison de l'ARN polymérase à l'ADN, 189-191
sporulation, 202, 204-205
transcription, initiation de la, 200-206
- Facteurs de relargage, synthèse protéique, terminaison, 120
eRF, 120
RF1, 120
RF2, 120
- Facteur stringent, 241
- FeSV, virus, oncogènes, 702
- Fibrinogène, 394
- Fibroblastes, normaux vs. transformés, 698-699
- Fixation, 407
- Flagelline, synthèse de la, 586-587
- Foyer cellulaire, 703-704
- Fragments d'okazaki, 317
- Fusion, température de, 58-59. Voir aussi Dénaturation
- β -Galactosidase, 36
opéron, induction d'un, 234
- Gamète, 19
ensemble haploïdes de chromosomes, 26
- Gène C, voir Gène, immunoglobuline C
- Gène du choc thermique, 201
chromatine, 544
transcription, facteurs de, 213
- Gène, copie d'origine d'un, 671-672
- Gène ovalbumine, ADNase I, 545-546
- Gènes :
abréviations, 36
actine, 394-395
activation des, 557-558
amplification, 625, 628-632
lignées instables, 629-630
lignées stables, 629
régions colorées de façon homogène, 629, 631
transfection, 633-634
- ARNt, 454-457
bactérien, 219
voir aussi Cistron
bg, 405-406
boîte, 428, 469
chevauchant, 89
codant pour une polyprotéine, 495-496
choc thermique, 201
facteurs de transcription, 213
chromosomes sur, 26-27
cl, répresseur, synthèse, 285
cII, 285-286
lysogénie vs. lyse, 291
cIII, 285
c-myc, 706-707
cob, 469
coconversion, 577
colinéarité procaryotique, 86-89
86-89
complémentants, 33
complémentation *cis-trans*, test de,
- Gènes (*suite*)
33-34, 389-391
c-onc, 702
concept, 93-94
constitutifs, 383
contenu en ADN, 411-412
contrôle coordonné, 220-223
contrôle métabolique, 31-32
copie liée à l'expression, (ELC), 672-673
copie non allélique, 400
copie d'origine, 671-672
cro, 275
antirépresseur, 288-289
cascade lytique, 289-290
transcription, 287
croisement en retour, technique de, 22
crossing-over inégal, 402-403
dam, 344
découverte *adn*, 319
découverte des, 18-22
définition, 3
délétion, 35-36
dhfr, 628-630
DHFR de mammifères, 387
disposition linéaire, 27-30
enzymes, association d', 31-32
erbB, 703
essentiels, 272
état diploïde, 19-20
état haploïde, 19-20
évolution homologue, 393
expression des gènes, 202
extrachromosomique, 512-513
extranucléaire, 422
familles :
définition, 399
protéines abondantes, 411-412
ftsZ, 300
ftsZ, 687
globe, structure interrompue, 387
hémizygote, 27
hsdRMS, 338-339
htpR, 201
immédiats, 262, 264, 272
immunoglobine C, 643
commutation de classe, 655-656
réaction de jonction, 646-648
recombinaison, 650-653
immunoglobuline V, 643
commutation de classe, 655-656
mutation somatique, 658-659
réaction de liaison, 646-648
recombinaison, 650-653
intégrateurs, 492-494

Gènes (suite)

interrompus, voir gènes interrompus,
isolement, 72-94
génom, 359-361
Kruppel, 245
lac, 36
lacI :
 mutation, 230
 mutations spontanées, 55
lactose, 220-223
lacZYA, 221-222
LDL, récepteurs aux, 395
lin-12, 695
loci, 29-30
 mutation, 369
lux de, 383
marqueur de restriction, 80-82
ménage, de 383
micF, régulateur, 245-247
mosaïque, 387
 complémentation *cis-trans*, 389-391
 gène, interruption de l'évolution
 d'un, 392
moyens, 272
N, 275
non essentiels, 272
onc, 615
oriC, 298
oriC, répllication, 325-326
orig, 327
oxi3, 428
 codant pour une polyprotéine, 494-495
POMC, 494-495
Q, 265
recA, recombinaison-réparation,
 347
recBC, 347-348
régulateur, 221
régulation, de, 221
répartition indépendante, 20-22
retardé, 262, 264-265, 272
rétroviral, 612-613
rho, 261
rnpA, 455
rnpB, 455
set, 493
Sgs, 552-553
SOS, 348-350
sous-méthylés, 549-550
structural, voir gène de structure
TcR, 659-660
td, 481
trans, agissant en, 34-35
transcription, 558

Gènes (suite)

contrôle de, 492-493
trns, agissant en 34-35
trp :
 atténuateur, 253-254
 séquence des peptides signals,
 253-258
 tumorigénicité *c-ras*, 704
uvr, 347-348
virulence, 677-678
v-onc, 702
5S, 418-419
Gènes, groupe de :
 alignement en demi registre, 441-442
 alignement en registre, 441-442
 ARNt, 454-457
 bactérien, 219
 codant des histones, 412-415
 contrôle coordonné, 220-223
 copies non alléliques, 400
 crossing-over inégal, 402-403
 définition, 399
 fixation du crossover, 420-421
 gènes des chaînes lourdes,
 assemblage des, chez la souris,
 648-649
 gènes de globine, 400-402
 gènes des immunoglobulines, 645
 homeo box, 696-697
 multiples copies actives, 419-421
 phages T7 et T4, 272-274
 réorganisation, 405-406
 tandem, 412-415
Gènes interrompus :
 caractéristiques, 386
 cartographie de restriction, 385-387
 codant pour l'ARNr et l'ARNt, 388
 découverte, 87-89
 épissage des ARN, 386
 épissage, profil d', 390-391
 évolution, 392-396
 exon-intron, fonctions, 388-389
 exons, 88
 introns, 88
 organisation, 385-396
Gènes, lactose, 220-223
Gène de structure, 220-223
 ADN non répétitif, 377-378
 batterie, 493
 définition, 386-387
 ensembles, 493
 familles, 399-400
 nodule activateur, 492-493
 site receveur, 492-493
 site sensible, 493
Gènes tumorigènes *c-ras*, 704

Génétique, cartographie :

 ADN chromosomique, séquences d',
 82-83
 chiasma, formation des, 29
 continuité, 30-31
 un gène, isolement d', 72-73
 groupe de liaison, 30-31
Génétique, marqueur, 36
 cartographie de restriction, 78-83
Génétique, matériel, 3
Génétique, polymorphisme, 78-79
Génétique, stabilité de l'organisation,
 589
Génique, expression :
 ARN, 183-184
 régulation, 244-245
 chaîne lourde, 656-657
 constitutive, 223
 contrôle d'inversion, recombinaison
 spécifique d'un site, 586-588
 déméthylation, 549-551
 écouvillon, 513-515
 gènes des, immunoglobulines, 641
 gènes eucaryotes dans les systèmes
 procaryotes, 364-366
 gènes non répétitifs, 378-379
 gènes précoces, 202
 gènes tardifs, 202, 272
 hn ARN, 490-491
 modèles de contrôle, 491-493
 niveau basal ou réprimé, 253
 organelle, 424-427
 phagés, gènes des, 202
 phage T4, cascade lytique du, 273-274
 protéine ribosomale, 242
 sous-méthylation, 549-550
 terminaison, 249
 transfection, 635
 variation de niveaux, 381-382
 voie alterne, 90
Génique, régulation, 681-697
 loci complexes, 689-695
Génome :
 ADN plasmide Ti, interaction, 675-680
 ADN répétitif, 374-375
 bactérien, 503-507
 boucles structurales, 505-506
 banque, 361-362
 chromosomes, 499-518
 composants des séquences, 371-372
 digestion des fragments de
 restriction, 360
 D. melanogaster, variation de la
 spécificité tissulaire, 626-628

Génome (*suite*)

- domaines, 505-507
- eucaryotique, 367
 - séquences répétitives, 372-373
- gènes des ARNm répétés et des ARNt, 415
- grande levure, 427-428
- nucléosomes, mise en phase des, 532-533
- haploïde, 368-369
- isolement de gènes, 359-361
- mitochondrial :
 - chez les mammifères, 428-430
 - évolution, 394-395
 - mutant vs. type sauvage, 35-36
 - opérer des changements, 625-636
 - phage lambda, sites *cos* du, 501
 - séquences amplifiées, 628-632
 - séquestré dans les organelles, 422-431
 - taille, 367-369
 - ADN non répétitif, 372
 - viral, 500-503
- Génotype, définition, 17
- α -Globine, groupe de gènes de l', 400-402
 - thalassémies α , 402-404
- β -Globine :
 - ADNase I, 545-546
 - divergences des séquences, 407-408
 - gènes, divergence des sites de remplacement dans l'évolution des, 409
 - groupes de gènes, 400-402
 - thalassémie, 404-405
 - pseudogènes, 411
 - sites hypersensibles, 553
 - thalassémies β , 404-405
- δ -Globine, divergence des séquences, 407-408
- Glucocorticoïdes, transcription, 215-216
- Glycoprotéine Prp, 92
- Glycoprotéine variable de surface (VSG), 670-675
 - site d'expression, 673-674
 - structure, 671-672
 - copie liée à l'expression, 673-675
- Glycoprotéines, 7
- GMP-PCP, synthèse protéique des composés, 115-116
- Globine, gènes de :
 - évolution, 405-406
 - évolution, interrompue, 393
 - évolution, horloge de l', 408-409
 - exons, structure des, 393

Globine (*suite*)

- organisation en groupes, 400-402
- G, protéine, 715
- Gradient de densité, 433
- Groupes phosphates, 41
 - riches en énergie, 44
- GT-AG, règle, 389
- GTP :
 - ribosome, fidélité de la traduction du, 157
 - synthèse protéique, facteur d'élongation, 115-116
 - translocation, 117-119
- GTPase, activité dépendante du ribosome, 110-111
- Guanine, 41
- Guanylyl transférase, 169
- Gyrase :
 - modèle de l'inversion du signe, 580-581
 - supertours négatifs de l'ADN, 580-582
- Hb Lepore, 404-405
- Hela, épissage des cellules, 474-475
- Hélice α , 9-10 *Voir aussi* ADN, double hélice
- Hémiméthylation, ADN satellite, 549
- Hémoglobine :
 - chaîne polypeptidique, 32
 - gènes, organisation de groupes de, 400-402
 - Hb anti-Lepore, 405
 - Hb Kenya, 405
 - Hb Lepore, 404-405
- Héparine, 187
- Hérédité :
 - chromosomes et, 22-26
 - cytoplasmique, 422
 - maternelle, 422-423
 - positive, 29
 - virale, 18
- Hétérochromatine :
 - ADN satellite, 434
 - constitutive, 508
 - centromère, 509-510
 - définition, 508
 - facultative, 508
 - section fine, 507-508
- Hétérozygotes, 20
- Histidine, contrôle de l'atténuation, 258-259
- Histone :
 - ADN associé au nucléosome, 539
 - chromatine, 520

Histone (*suite*)

- cœur, 521-522
- gènes, groupes de, 412-415
- modification transitoire, 546-549
- octamère, 525
 - conservation, 539-540
- riche en arginine, 520
- riche en lysine :
 - légèrement, 520
 - très, 520
- HML, commutation du locus, 605-607
- HMR, commutation du locus, 665-667
- hn ARN, 483
 - ARNm dérivé de, 485-487
 - chez les mammifères, 490-491
 - contrôle post-transcriptionnel, 489-491
 - distribution de taille, 484
 - in vitro*, 488
 - polyadénylation, 485-487
 - extrémités 3, 487-489
 - taille et stabilité, 483-485
- hnRNP, 485
- HO endonucléase, 669-670
- Hogness, boîte, 208-209
- Holoenzyme :
 - ARN polymérase bactérienne, 185-188
 - Bacillus subtilis*, sensibilité à, 2306
 - complexe ouvert, 190
 - complexe fermé, 189-190
 - facteur, 190
 - liaison étroite, 190
 - NusA, protéine, 267
 - site de liaison, perte du, 190
- Homeo box, 695-697
- Homologues, 23
- Homomultimères, 32
- Homozygotes, 20-22
- Hpal, enzyme, 78
- HPHF, 405
- HSR, régions colorées de façon homogène, 629, 631
- Hybridation :
 - acide nucléique : ADN, clivage de l', 78
 - appariement des bases, 58-60
 - ARN, épissage de l', 473
 - ARNm :
 - non répétitif, 379
 - traceur, 378
 - ARNm $\times$ ADNc
 - complexité des ARNm, 380
 - Rot, paramètres, 381
 - colonie, 362-364
 - cytologique, voir *in situ*

Hybridation (*suite*)

- filtre, 59-60
- gène, isolement d'un, 361
- gènes de structure, cinétique, 377-378
- in situ*, 434
 - bandes sur un chromosome, 516-517
 - dysgénésie hybride, 622-624
 - transfection, 633-634
- liquide, 59, 370-371
- marche sur le chromosome, 363-364
- saturation, expériences de, 379

Hybridome, 658

Hydrogène, liaison :

- ARNt, structure tertiaire des, 125-126
- bases azotées, 45-47
- définition, 8

Hydroxyméthylcytosine, 329-330

Hyperchromicité, 58

Hypothèse du déclenchement par la membrane, 178-179

IF2.fMET-fARNt, 110

Immortalisation, 699, 711-715

Immune, réponse, 641

Immunité :

- effet, 270
- région, 278

Immunoglobuline :

- apparentée au récepteur des cellules, 659-660
- chaîne légère (L), 642-645
 - kappa, jonction V-J, 646-649
 - kappa, 643
 - lambda, recombinaison V-J, 645-645, 648
 - lambda, 643
- chaînes lourdes (H), 642-645
- exclusion allélique, 653-655
- gènes, assemblage des, 641-645
- réactions de jonction, 646-647
 - diversité, 649
- région constante (C), 642-643
- région hypervariable (HV), 643
- région ossature (framework (FR) region), 643
- région variable (V), 642-643
- segment D, 646-647
 - réaction de jonction, 649
- segment J, 646-647
- synthèse, 656-657
- translocation, 708

Indice d'enroulement topologique, 70

Inducteur :

- définition, 220
- gratuit, 220
- opéron *lac*, 222-223
- protéine répresseur, 231-233

Induction :

- définition, 220
- opérons, 234

Information de position, 94

Inhibition multicopie, 595

Inhibition par le produit, 252-253

Initiation :

- ADN, synthèse d', 320-322
- ARNm polycistronique, 163-165
- ARN polymérase-promoteur, interaction, 183-218
- cellule cancéreuse, 711
- codon, 109
- complexe de pré-initiation, 217
- eucaryotique, 112-114
- migration des sous-unités, 172-173

facteurs (IF) :

- eIF2, 112-114
 - sous-unités, 113-114
- eIF2B, 113-114
- eIF3, 112-113
 - sites actifs du ribosome, 154
- recombinaison, 568-569
- réplication, 295, 300
 - ADN, double-brin d', 326-327
- site de liaison au complexe ribosomal, 106, 108-109
- sites du nucléotide, 186
- synthèse protéique, 105, 108
- transcription :
 - catalyse par l'ARN polymérase, 184-185
 - dépendance des protéines, 217
 - « enhancers » bidirectionnels, 213-216
 - facteur sigma, 200-206
 - facteurs, 213

Insertion :

- définition, 35
 - locus *white*, 682-683
 - modules, 592-594
 - mutations ponctuelles, 53
 - oncogène, 706-711
 - séquences : transposons, 590-592
- Insuline, interruption de l'évolution du gène, 392-394

Intégration, ADN de lambda, 582

Intéactions hydrophobes, 9

Intéactions ioniques, définition, 8

Interactions protéine-protéine, 284-285

Intercalaire transcrit, 416-418

Intermédiaires de préamorce :

- mouvement du primosome, 322-325
- réplication, 325-326

Interphase dans le cycle cellulaire, 23

Intron :

- ARN nucléaire ovomucoïde, 472-473
- auto-épissage, 465-468
- cartographie des mutations, 682-685
- classe I : auto-épissage, 466-467
- intermédiaires en lasso, 477-478
- maturase, 470
- classe II : intermédiaire en lasso, 477-478
- maturase, 470
- réactions apparentées, 481
- codant pour les ARNr et les ARNt, 388
- codant des protéines, 468-472
- complémentation des exons, 389-391
- définition, 88

- éléments Ty, 619-620
- évolution des gènes interrompus, 385-387, 392

mitochondrial, 465-468

phosphoesters, réaction de transfert des, 463-464

pseudogènes, 410

réaction linéaire, génération des dimères, 464

séquence consensus, 388-389

Inversion :

- duplicative, 598-599
- modèle, réaction de jonction, 651-652
- ocelliless*, 627-628
- recombinaison spécifique de site, 586-588

Irradiation UV, détérioration de l'ADN, 347

IS, voir Insertion, séquences,

Isomères topologiques, 578-579

Isopropylthiogalactoside (IPTG), 220

Isoschizomères, 549

Kasugamycine, résistance à la mutation des ribosomes, 153

Kinétochore, 510

Kirmomycine et inhibition de la synthèse protéique, 116

Lamarckien, processus, 644

LDL, gène du récepteur au, 395

- Leader peptidase, 178-179
 Lecture au-delà :
 ARNt suppresseur, 141
 définition, 249
 Léghémoglobine, 393
 Leucémie myéloïde chronique, 708-709
 Levure :
 éléments Ty, 618-620
 épissage des ARNt, 459-461
 génomomes des grandes mitochondries, 427-428
 épissage, 470-471
 mutations de la lignée W, 470-472
 replicons, 303
 séquences activatrices en amont, 215
 type sexuel :
 communication, 665-666
 loci silencieux et actifs, 665-666
 modèle des cassettes :
 actif et silencieux, 666
 séquences, 667-669
 substitution, 670
 Liaison croisée, centres actifs du ribosome, 156
 Liaison, nombre de :
 ADN, topologie de l', 578-579
 enroulement, 70
 nombre de torsion, 70
 superenroulement, 69-70
 Liaison phosphodiester :
 ARN, 445-446
 enzyme cœur de la synthèse d'ARN, 187
 Liaison, polymorphisme de restriction, gènes, 72
 groupe de, 30
 Ligation à bouts francs, 357-358
 Ligation, réaction de, épissage des ARNt, 459-460
 Lignée germinale :
 ADN transfecté, 634-636
 information, diversité de l', 647-649
 profil, immunoglobuline, 641-645
 promoteur, fonction du, 208
 Lipides, 4
 Loci :
 cellules T, récepteur des, 659-660
 complexes, régulation des gènes, 689-695
 définition, 29-30
 histocompatibilité, 660-663
 HML, 665-667
 HMR, 665-667
 létaux, 518
 MAT, 664-666
 Loci (*suite*)
 Shrunken (Sh), 608
 Waxy (Wx), 608
 white, 682-685
 voir aussi Loci complexes
 Loci complexe :
 achaete-scute, 689-690
 BX-C, 689-690, 694-695
 Notch, 690, 694-695
 Loci, MAT, 664-666
 transposition unidirectionnelle, 669-670
 Lymphocytes : ARN, maturation des, 657
 cellules B, 641-642
 tumeurs, 707
 cellules T, 641-642
 Lyse, rapide, 31
 vs. lysogénie, 289-291
 Lysogénie :
 cascade, 287-288
 circuit autogène, 276-279
 définition, 269-270
 phage lambda, 582-583
 recombinaison spécifique de site, 582
 vs. lyse, 489-291
 Lytique, cascade : phage T4, 273-274
 cascade de contrôle, 270-272
 phage lambda, 274-276
 Lytique, état : phage lambda, 582-583
 recombinaison spécifique de site, 582
 Lytique, infection :
 antirépresseur, 287-289
 définition, 202, 269-270
 précoce, 270
 retardée, 270
 Macromolécule :
 définition, 4
 expression des gènes, 219-220
 hérédité et, 3-4
 liaison hydrogène, 8
 Maladie de la gale du colet, 675
 Maladie HbH, 402
 Marquage des extrémités, technique de, l'ADNase, sites hypersensibles à, 552
 coupure, 527-528
 restriction, carte de, 77-78
 voir aussi Technique indirecte de marquage des extrémités
 Marquage indirect des extrémités, technique de, 532, 555
 Marquage radioactif : hnARN, 486-487
 réplication, 297-298
 Marqueurs de restriction :
 diagnostic des maladies, 80-82
 polymorphisme, 79-81
 carte de lésion, 82
 Méiose, 23-26
 ADN, double brin hétérologue d', 577-578
 chromomère, 513
 chromosomes en écouvillon, 513-517
 chromosomes, partage des, 23, 25
 complexe synaptonéal, 564-565
 conversion génique, 420, 577-578
 Membrane plasmique, 3
 Mendel :
 Deuxième loi, 20-22
 déviation, 27
 Première loi, 19-20
 Messenger subgénomique, 612-613
 Métaboliques, contrôle génique des voies, 31-32
 Métabolisme, erreurs innées du, 17
 Métastase, 699
 Méthotrexate, 629
 Méthylation :
 acides aminés, 7
 ADN, sauvegarde de l', 336-338
 ARNr, clivage des, 450-451
 dam, système, 344, 595
 enzyme de restriction type III, 341-342
 histones, modification des, 546-549
 nucléosomes, assemblage des, 547
 voir aussi Déméthylation, Expression des gènes ;
 Micrococcale, nucléase, 521-523
 ADN nucléosomale, 524
 chromatine, digestion de la, 543-544
 clivage, 531-533
 échelle, 544
 sites hypersensibles, 553-554
 Micro-hétérogénéité, 414-415
 Migration de la fourche, 565-568
 simple-brin, 566-568
 Minichromosome, 530-531
 Mitochondrie :
 ARN polymérase, 189
 codons, 105
 complexes protéiques dans la synthèse protéique, 425-426
 ensemble d'ARNt, 137
 évolution des gènes, 394-395
 expression des gènes de l'organelle, 424-427
 variations d'ADN, 423-424

- Mitose :**
 constitution du chromosome, 23-24
 cycle cellulaire, 23-24
- Modèle de la correction intempestive,**
 420-421
- Modèle de délétion, réaction de liaison,**
 651-652
- Modification :**
 base ARNt, contrôle de codon, 135-136
 enzymes, 335-336
 opération, 336-338
- Molécule :**
 classification, 4
 concatémériques, 501-502
 hybride ARNm-ADN, 358-359
 liée, 565
 liée, échange de brins catalysé par RecA, 574-575
 résolution, 566
 ouverte, 68-69
 petite, 4
 processing, 445
 relâchée, 68-69
voir aussi Macro molécule
- Molécule d'ADN double brin**
 hétérologue (hybride), 565
- Molécules extrachromosomiques,** 512-513
- Molécule de jonction,** 565
 RecA, échange de brin catalysé par, 574-575
 segment J, 646
- Mouvement au hasard et fixation,** 407
- Mouvement du primosome,** 322-325
- mRNP,** 165-166
- MspI, enzyme,** 549-550
- Mud, phage,** 366
- Multimère hétérogène,** 32
- Mutagènes,** 52-53
- Mutant, définition. Voir aussi Mutation**
- Mutation :**
 acridine, 51-52
- ADN :**
 arrêt et retard, 331
 dommages, 345-347
 incorporation, 53-54
- ADN polymérase,** 316
- antisens, lysogénie,** 287
- ARN I,** 309-310
- ARNt :**
 codons, 137-140
 reconnaissance, 128
- ARNt, mutants,** 137-139
- arrêt lent,** 319
- réplication,** 325-326
- Mutation (suite)**
 arrêt rapide, 319
 atténuation, 255-256
 BX-C, 688-690
 domaine infra-abdominal, 692-693
 cadre de lecture, 51
 cartographie des délétions, 35-36
 cartographie génétique, 73
 cartographie, niveau moléculaire, 35-36
 chi, 584-585
 chromosomes, nodulisation des, 517
cis, agissant en, 517
 codon, dégénérescence du, 104-105
 complémentation *cis-trans*, 33-34
 complémentation, groupe de, 33-34
 concentration des points chauds, 53
 conditions létales, 35-36
 conditions non permissives, 36
 conditions permissives, 36
 constitutive, 223-224
 délétion atténuateur, 258
cis — dominant, 225-226
 contrôle positif, 284-285
 cy, 286
 cytochrome b, 469
 décalage du cadre, 51-52
 définition, 17
dna, 319
 "down", séquence consensus, 194
 dysgénésie hybride, 623-624
 exon, 682-685
 fixée, 407
 gène *lac I*, liaison à l'ADN, 230
 gènes, transcription des, 202
 "high-hopper", 594-595
 homéotique, 688-689
 hypothèse du signal, 176, 178-179
 induite, 52
 intron, 88, 468
 auto-épissage, 468
 cartographie, 682-685
 introns agissant en *trans*, 469
 jonctions d'épissage, séquences consensus, 476-477
lac ZYA, 221
 "leaky", 53
 loci *MAT*, 664-666
 neutre, 407
 substitutions, 56
 non sens :
 supresseurs ARNt, compétition entre codon, 140-141
 138-139
 terminaison de la synthèse protéique, 120
- Mutation (suite)**
 traduction des ARNm, 164
ocelliless, 627-628
 opéron, non inductible, 226-227
 petite :
 ADN mitochondrial de levure, 430-431
 neutre, 431
 nucléaire, 431
 suppressive, 431
 phage T4, 33-34
 polyoma, 554
 ponctuel :
 cartographie de restriction, 78-79
 constitutive, 228-229
 définition, 35
 séquence consensus, 194
 simple appariement de bases, 52-53
 terminaison, 249
 transition, 52
 transversion, 52
 promoteur, 198-200
 protéine répresseur, 223
 proto-oncogènes ras, 703-706
 récessive, 32, 695
 recombinaison, 31
 relâché, 241
 rétrorégulation *sib*, 449-450
 réversion :
 second-site, 56
 vraie, 56
rho, 261
sad, 151
 segmentation, 685-687
 interruption, 686
 profil, 686
 règle de la paire, 686
 séquence cœur, 583-584
 séquence consensus, 194
 silencieux, 56, 407
 somatique, 658-659
 sous-unité ribosomale sensible au froid, 151
 spontanée, 52
 gène *lacI*, 55
suA, 261
 supresseur, 51
 supresseurs ARNt, 138-140
 terminaison de la synthèse protéique, 120
 terminologie, 36
tnpA, transposition, 602
 traduction, fidélité de la, 157
 type sauvage, 36
 ultrabithorax, 690-692

Mutation (*suite*)

- virulent, 276
- protéine répresseurs, 285
- vitesse, 55-56

Mycoplasme :

- codons, 105
- réplication, 91-92

Myoglobine, 393

Neurospora crassa, 31-32

- cob1, intron, 467
- gènes interrompus, 388
- génome mitochondrial, 425-427

N-formyl-méthionyl-ARNt, voir ARNt Met

Nick translation, 316

N oligosaccharides liés à l', 7

Northern blotting, 361

Notophthalmus viridescens,

- chromosomes en écouvillons, 514

Noyau, 3

- matrice du, 485

Nucléase, 61

Nucléotide :

- coupure, 505-507
- définition, 499
- génome bactérien, 503-507

Nucléole, 418

Nucléoplasmine, 541-542

Nucléoside, définition, 42

Nucléosome :

- assemblage, 539-540
- protéine non histone, 540-542
- chromatine, cheminement de la fibre de, 533-535
- chromatinienne, sous-unité, 521-524
- condensation, rapport de, 542-543
- échelles, 522-524
- forme, 525
- gènes transcrits, 542-543
- histone :

- tétramère H3₂ H4₂, *in vitro*, 540
- 529-530

lysé, 521

- mise en phase, séquençage de l'ADN, 531-533

- modèle symétrique, 529

- multimère, 523-524

- particule cœur, 525-526

- Tetrahymena*, ADNr de, 554-555

Nucléotide :

- ARNt modifié, 133-134
- classification, 4
- définition, 42
- gènes morcelés, séquences, 386

Nucléotide (*suite*)

- porteur de phosphate, 43-44
- ppGpp, 241
- "pulse-chase", expérience de, 486-487

Nus, loci, 267-268

- Okazaki, fragment d', ARN, amorçage de l', 317-318
- primosome, mouvement du, 322-325
- réplication de double-brins d'ADN, 325-329

Oligosaccharides, 7

Oncogène(s), 698-715

- activité de tyrosine kinase, 714-715
- amplification dans les tumeurs 710
- c-myc*, 708-711
- interaction avec les proto-oncogènes, 715

L-myc, 711

- mutation, 703-704

myc, 706-711*N-myc*, 711

- rétroviral, 701-703

- transfection, test de, 704, 711-712

- transformation virale, 699-701

- virale, 714-715

Oncogénèse, 615

Oncoprotéine, 712-715

Opérateur, définition, 223-225

Opéron :

- atténuation, 252-256
- circuits, positif et négatif, 235-236
- cis*, agissant en, 225-227
- contrôle autogène, 234-235
- définition, 221-222
- déréprimé, 236
- diploïdie partielle, 223
- E. Coli*, 224-225
- histidine, 258-259
- ilv*, 258
- inductibilité, 223-224, 236
- induction, paradoxe d', 234
- lac*, 222-223

- CAP, site de liaison à la coiffe, 239-240

L11, 242-243

- opérateurs dual, 557

- organisation, 242-243

- protéine répresseur :

- ARNr, 451-452

- contacts, 228-229

- dimère, 278-285

- sites de liaison, 282-283

- répressible, 236

Opéron (*suite*)

- répression multivalente, 258

- rif*, 242-243

- rrn*, 451-452

- spc*, 242-243

- str*, 242-243

- sur-réprimé, 236

- S10*, 242-243

- thr*, 258

- trp*, 234-235

- transcription, 237-238

- tryptophane, contrôle atténuation du, 252-253

- tyrU*, 454

Opines, 675-676

Organelle :

- ADN, recombinaison d', 430

- gènes, expression des, 424-427

- du génome, séquestration du, 422-431

- hérédité non Mendélienne, 422-423

- modules d'ADN circulaires, 423-424

- ribosome, 145-146

Organisateurs nucléaires, 418

Organisme :

- diploïde, 20

- hétérozygote, 20

- homozygote, 20

- multicellulaire, 4

- unicellulaire, 4

Organisme de type sauvage, définition, 17

Oscillation, hypothèse de l' (Wobbe hypothesis) :

- modification des bases, 135-136

Palindrome, 63-65

- Palindrome, axe de symétrie, 63

- modes de terminaison, 249-250

- symétrie par rapport à un point, 63

Paranémique, jonction, 574-575

Parasegments, 685

Particule cœur :

- conservation, 525-526

- structure du cristal, 528-529

Pathogénicité subvirale, 91

Pentose, 41

- désoxyribose, 42

- ribose, 42

Peptide :

- définition, 5

- liaison, 4-5

- translocation, 116-120

Peptidyl ARNt, 106

- synthèse protéique, initiation de la,

- Peptidyl ARNt (*suite*)
 110
 Peptidyl transférase, 116-117
 activité, 156
 Peptide leader, liaison du ribosome au,
 253-257
 P, facteurs, 622-623
 Phage :
 ADN, sauvegarde de l', 336
 chimérique, 353
 définition, 18
 échanges génétiques, 72
 génom, 270-272
 hybride, 353
 infection lytique, 270-272
 morphogénèse, 329
 vecteurs de clonage, 355
 Phage lambda :
 antiterminaison, 262-264
 capside, assemblage de la, 355
 culture du type sauvage, 277-278
 gènes précoces, 264-265
 liant le dimère répresseur, 556
 lytique :
 antiterminaison, 274-276
 circuit autogène, 276-278
 vs; état lysogénique, 582-583
 région régulatrice, 277-278
 réplication bidirectionnelle, 501-503
 unités transcriptionnelles, 264-265
 vecteur de clonage, 355
 Phage Mu, 366
 segment G ou réversible, 587-588
 Phage SPO1 transcription, 202-203
 Phage SP6, 364
 Phage T2 :
 ADN comme matériel génétique, 39-40
 E. coli, infection d', 39
 Phage T4 :
 capside, assemblage de la, 501-503
 carte génétique, 31
 gènes, groupement de, 272-274
 mutations, groupe de
 complémentation, 33-34
 régulation autogène, 244
 réplication, 329-331
 Phage T7 :
 ARNm précoces, 448-450
 groupe de gènes, 272-274
 réplication, 329-331
 Phage ϕ X174, 322-325
 ADN non enroulé, 327-328
 séquences exogènes, 633
 Phénomène d'effet de position, 628
 Phénotype :
 défini, 17
 parental, 22
 polymorphisme, 78-79
 marqueurs de restriction, 80-82
 réciproque, 22
 recombinant, 22
 tumorigène, 704
 Phénylcétonurie, 17-18
 Pheromones, 665
 Phosphoprotéine, 7
 Phosphorylation :
 acides aminés, 7
 histone, 547
 sérine, 547
 Photoréactivation, 342-343
Physarum polycephalum, 466
 gènes morcelés, 388
 Plaque, lysogénie, 276-278
 Plasmide :
 ADN transfecté, 634-635
 bactérien, transposons, 590
 chimérique, 353
 ColEI, 308-309
 défini, 223, 269
 fragments CEN, 510-511
 hybride, 353
 incompatibilité, 307-311
 levure, 501-511
 linéaire, 464
 nombre de copies, 307-311
 primosome, 322-325
 réplicon, 298
 résistance aux antibiotiques, 354
 sauvetage de l'ADN, 336-337
 séquence télomérique, 51
 Ti, 675-680
 agropine, 676
 fonctions, 679-680
 gènes des plantes et des bactéries,
 676
 nopaline, 676
 octopine, 676
 répétitions identiques, 678
 Ri, 676
 type sexuel, 666-667
 vecteur, 356
 vecteur de clonage, 354
Pneumococcus, types, 38-39
 Point de départ :
 ARN polymérase II, promoteurs
 amont, 206-209
 définition, 185
 influence sur l'initiation de la
 transcription, fréquence, 217
 unité transcriptionnelle amont et aval,

 Point de départ (*suite*)
 192-193
 Points chauds, définition (Hot spots),
 55
 Polarité :
 ARNm polycistronique, 164
 facteur rho, 260-261
 Poliovirus :
 ARN, 332-334
 traduction des ARNm, 173
 Polyadénylation :
 ARNm eucaryote, 168-169
 épissage des ARN nucléaires, 473
 expression forte des gènes, 656-657
 hnARN, 485-487
 extrémités 3', 487-489
 in vitro, 488
 poly(A) —, 168
 poly(A) +, 168
 queue poly (A), 168
 terminaison, 251
 transcriptase inverse, 168
 transcrit primaire, 492
 Polymorphisme :
 ARNr et ARNt, 395
 divergence de séquence, 407
 génétique, 78-81
 loci d'histocompatibilité, 660-661
 Polynucléotide :
 chaîne, 42-43
 antiparallèle, 47
 ARN polymérase-promoteur
 complexes, 198
 synthétique, traduction de *E. Coli*,
 103-105
 Polyprotéines cellulaires, 494-496
 Polyribosome, voir Polysome,
 Polysaccharide, 4
 Polysome :
 cytosquelette, 173-174
 défini, 101
 libre, 173
 lié à la membrane, 173
 synthèse protéique, 101-102
 taille, 101
 Préproinsuline, 174
 Préprotéine, 174
 Primase, 322
 Primosome DnaB, mouvement du, 323-325
 Prion, 92-93
 Procaryotes :
 appareil de réplication, 319-320
 ARNm, 159
 colinéarité, 86-89

Procaryotes (*suite*)

- défini, 3
- expression des gènes eucaryotes, 364-366
- synapse de recombinaison, 564
- topoisomérase, 579-581
- Proinsuline, 174
- Promoteur :
 - actif, hypersensibilité à l'ADNase, 551-555
 - ARN polymérase :
 - interactions, 183-218
 - liaison *in vitro*, 195-200
 - ARN polymérase II, 206-209
 - anomalie de taille, 210
 - composants de séquence, 210-211
 - structure multipartite, 209-211
 - ARN polymérase III :
 - en aval, 216-218
 - localisation interne, 217
 - ARNt, 216-217
 - balayage de « linker », 208
 - boîte CAAT, 211
 - boîte TATA, 208-209
 - composants, 195
 - défini, 185, 214
 - enhancer :
 - bidirectionnel, 213-216
 - cellulaire, 214
 - répétition de 72 pb, 214
 - espace intercalaire non transcrit, 416-417
 - facteur de contrôle, 217
 - gène morcelé, 389-390
 - gène V, 650-653
 - liaisons, 207-208
 - ARN polymérase III, 216-217
 - systèmes *in vitro*, 208-211
 - liant la CAP, 239-240
 - locus *white*, 684-685
 - mutations, 198-200
 - contacts entre opérons, 228-228
 - non inductible, 226-227
 - opérons *rrn*, 451-452
 - P_{RE}, 285-286
 - réaction de jonction, 652-653
 - régulation des gènes, 681
 - relation entre opérons, 223-225
 - répression catabolite, 236-240
 - séquences consensus, 193-195
 - boîte CAAT, 211
 - boîte GC, 211
 - boîte « Pribnow », 194
 - boîte TATA (Hogness), 208-209
 - conservé, 194
 - facteur sigma de *Bacillus subtilis*,

Promoteur (*suite*)

- 205-206
- fonctions, 193-195
- reconnaissance du facteur sigma, 200-201
- séquence 10, 35, 194
- TATAAT, 194
- superenroulement et, 200
- synthèse de répresseur, 285-287
- système test :
 - in vitro*, 206-209
 - in vivo*, 206-209
 - séquences consensus, 209-211
- oocyte, 206-209
- transcription des ARN 5S, 216
- transcription, IS10R, 594-595
- unités de transcription, 192-193
- Promotion, cellule tumorigène, 711
- Proopiomélanocortine (POMC), 494-495
- Prophage, 269
 - intégration, 269
- Protéine, 504-505
 - acides aminés, 5-6
 - activité peptidyl-transférase, 156
 - allostérique, 12-13
 - antiterminaison, 265
 - apo-inducteur, 236
 - assemblage, 271-272, 503, 525
 - assemblage des nucléosomes, 541
 - auxiliaire, 330-331
 - chaînes, 4-7
 - classification, 4
 - cofacteur, 12
 - colinéarité, 86-89
 - complément, 661-662
 - domaine externe, 661-662
 - conformation, 11
 - forces non covalentes, 7-9
 - repliement, 11-13
 - conservée, 407
 - Cro, 275, 288-290
 - TFIIIA, 217
 - cytoplasmique, 422
 - dans la chromatine, 520-521
 - définition, 5
 - dénaturé, 11
 - divergence, 407
 - diversité structurale, 9-11
 - DnaA, 325-326
 - domaines, 12
 - engrailed, 688
 - homeo box, 696
 - évolution, 395-396
 - des gènes interrompus, 392-393
 - exons et, 395-396

Protéine (*suite*)

- facteur stringant, 241
- famille de gènes, 411-412
- G, 715
- gène codant, 31-32
- globulaire, 10
- glycosylation, 7
- groupe de haute mobilité (HMG), 520
- HLP1, 505
- hybrides vs bona fide, 366
- IHF, recombinaison, 584-585
- inhibition par rétrocontrôle, 13
- interaction coopérative, 557
- Int, recombinaison, 584-585
- intron codant, 468-472
- LexA, 348-349
- liaison à la coiffe, ARNm eucaryote, 172-173
- liaison simple-brin, 320-322
- se liant à l'ADN :
 - domaines, 280
 - E. coli*, 504-505
- modification, 4-7
- multimérique, 10
- non histone :
 - assemblage de nucléosomes, 540-542
 - chromatine, 540
 - nucléosome, 526
 - oncogénèse, fusion, 710
- organisation, immunoglobuline, 642-645
- repliement, 11-13
- résistance aux antibiotiques, 153
- Protéine pN, 265-268
- RecA, réponse SOS, 349-350
- Rep, 324-325
- Rep, ADN désenroulé, 328-329
- réplication, appareil du phage TH, 330
- répresseur, liant opérateur, 223-225
 - transcription, 227-228
- résistance aux antibiotiques, 153
- ribosomale :
 - expression des gènes, 242
 - interaction ARNr, 147-150
- secrétée, 174
- séquençage d'ADN, codant des multiprotéines, 89-90
- non homologue, 89
- séquentiel, 11-13
- site, actif, 11
- sous-unités, 10
- split, ribosome subunit, 150
- structure en feuillet plissé, 9-11
- structure de l'hélice α , 9-11

- Protéine pN (suite)**
 structure quaternaire, 10
 structure tertiaire, 9-10
 variable :
 glycoprotéine de surface, 670-672
 majeur, 675
- Protéine répresseur :**
 accumulation en surplus, 231-233
 ADN, liaison à l', 227-228
 déplacement site à site, 233-234
 multimérique, 229-231
 sites d'affinité, 232-233
 bras ou protusion, 231
 charnière, 231
 cœur résistant à la trypsine, 230
 contact ADN, 281-282
 contrôle allostérique, 223
 définition, 221
 déplacement, 232-233
 IPTG complexe, 220, 227, 231
 loci multiples, 234-235
 mutations, 223
 opérateur, contacts avec l', 228-229
 opérateur, liaison à l', 223-225
 phage lambda, 280-283
 arrangement en double hélice, 281-282
 pièces maîtresses, 230
 structure dimérique, 278-280
 synthèse, 285-287
 transcription, blocage de la, 227-228
 trp, 234-236
- Protéine ribosomale :**
 ARNr, 243-244
 EF-Tu, 242-244
 L7IL 12, 242-244
- Protéique, synthèse :**
 et antibiotique, 152-153
 appareil mitochondrial, 425-427
 ARNm polycistroniques, traduction des, 163-165
 ARNm, traduction ribosomale, 99-103
 ARNr 16S, 153
 ARNt comme adaptateur, 98-99
 codon de terminaison, 103-105
 cytoplasmique, 422
E. coli, composants bactérien, 101, 103
 élongation, 105
 complexe binaire, 114
 facteur T, 114-116
 facteur d'antiassociation, 108
 facteur de dissociation, 108
 facteurs de libération, 120
 eRF, 120
- Protéique, synthèse (suite)**
 RF1, 120
 RF2, 120
 facteur de recyclage, 111
 fidélité de traduction, 156-157
 hypothèse du déclenchement par la membrane, 178-179
 inhibiteurs, 116
 initiation, 105
 complexe ternaire, 112-113
 eIF2, 112-114
 eIF2B, 113-114
 eIF3, 112-113
 eucaryotique, 112-114
 facteurs, 108-109
 IF1, 108
 IF2, 108
 IF3, 108
 sous-unités 30S, 106, 108-109
 lignée d'assemblage, 97-120
 localisation cellulaire, 173-179
 polypeptidique, chaîne, 99-101
 ARNt initiateur, 109-111
 polysomes, 101-102
 pré et pro-séquences, 174
 réaction d'élimination, 111
 réponse forte, 240-241
 ribosome, site d'action du, 105-109
 A, site d'entrée, 106-107
 P, site donneur, 106-107
 séquence « leader », 174-175
 signal, hypothèse du, 175-179
 séquences d'ancrage, 176, 178
 signal de demi-transfert, 176, 178-179
 signal, particule de reconnaissance, 175-177
 terminaison, 105
 codons de, 120
 facteurs de libération, 120
 prématurée, 120
 translocation, 116-120
 transport cotraductionnel, 175
 transposons, 594
- Proto-oncogène, 702**
 activation, événements d', 706-711
 interaction oncogénique, 715
 mécanismes, 712-715
 ras, 703-706
- Provirus, endogène, 610-611**
- Pseudogène :**
 antigènes de transplantation, 662
 comme culs-de-sac de l'évolution, 409-411
 définition, 400
 gènes 5S et, 418-419
- Pseudogène (suite)**
 groupes, 400-401
 mûré, 410
 transcription de l'ARN inverse, 616-617
- Psoralène, 506**
 « Puffing » (nodulisation) :
 ADN, variation selon la spécificité tissulaire, 626
 chromosome polytène, 517-518
 « Pulse-chase », résultat du : hnARN, 486-487
 épissage, voies d', 473
 Purines, 41-42
 Puromycine, translocation, 117-119
 Pyrimidines, 41-42
- Ras, gènes, 703-706**
 Réaction au ralenti, 241
 Réaction de transfert phosphodiester
 épissage, 462-463
- Réassociation, cinétiques de, 369-372**
 ADN répétitifs, 375
 cinétique, pureté, 371
 composants de séquence, 371-372
 Cot, vitesse, 370-371
 hybridation, 377, 378
 intermédiaire, 371
 lents, 371
 rapides, 371
- RecA :**
 assimilation des simples brins, 572-573
 facilité d'échange de brin, 571-573
 recombinaison, conditions de, 574-576
- RecBC, endonucléase, 575-576**
- Recombinaison, 563-588**
 ADN apparu en double-brin, 565-566
 ADN des organelles, 430
 commutation de classe, 655-656
 conservative vs. répllicative, 597-598
 délétion, modèle de, 651-652
 distance sur la carte, 28
 double, 29
 double-brins, initiation de la cassure des, 568-569
 éléments Ty, 619-620
 généralisée, 563
 gènes V et C, 650-653
 génétique, 29
 interallélique, 576-578
 intermédiaires, isolement des, 569-571

Recombinaison (*suite*)

- modèle du croisement inégal, 402-403, 442
- modèle d'inversion, 651-652
- nodules, 565
- procaryotes vs. eucaryotes, 72-73
- RecA, échange de brins facilité par, 571-573
- réciroque :
 - cercle dimérique, 569-570
 - sites homologues, 569-570
- répliative, 563
- restriction, polymorphisme de, 79-81
- séquence cœur, 582-583
 - bras, 582-583
 - cassure et réunion, 582-583
- site spécifique, 563
 - cassure en chicane et réunion, 583-586
 - expression d'un gène, contrôle inversé de l', 586-588
 - reconnaissance, 582-583
- somatique, 643-655
 - exclusion allélique, 653-655
- spécialisé, 582-583
- synapse entre double-brins d'ADN homologues, 564-565

Recombinaison-réparation, enzymes de, 344-345

système de, chez *E. Coli*, 347-348

Reconnaissance, hélice de, 281

Redondance terminale, 502

Régulateurs positifs, 236-240

Régulation coordonnée, ensemble de gènes, 220-223

Renaturation :

- ADN, 57-58
- ADN non répétitif, 375-377
- ADN répétitif, 375-377
- définition, 58
- réassociation, cinétique de, 370-371

Réparation post-répliative, 344

Répétitions :

- directes, 590-591
 - T-ADN, 678
- extrémités inversées, transposons, 590-592
- fréquence de, 363
- hiérarchisées, 435-438
- identique, 435, 442
- longues terminaisons, 612-614
- variation des unités, 439-440
- 72 pb, 214

Répétitions en tandem :

- ARNr, gènes des, 416-418
- gènes, groupes de, 412

Répétitions en tandem (*suite*)

- nucléosomes, mise en phase des, 533
- télomère, 511-512
- variations de longueur et d'espèces, 416-417
- Répétitions identiques, satellite, 435
- Répétitions inversées, 63-65
- Répétitions organisées, 435-438
- Répétitions terminales inversées, transposons, 590-592
- Répétitions terminales longues, 612-614
 - oncogènes, 706-707

Réplicon :

- chromosome eucaryote, 300-303
- définition, 295
- eucaryote vs. procaryote, 302
- fusion, 598
- génomme bactérien, 298-299
- levure, 303
- linéaire, 331-334
- origine, 295
- réplication, comme unité de, 295-311
- structure θ , 297, 303
- taille, 301-302
- terminal, 295

Réplication :

- acides nucléiques, appariement des bases, 49-50, 90-93, 314
- activité d'initiation, 313
- ADN de l'adénovirus, 332-334
- ADN colE1, 308-309
- ADN double-brin, 325-329
 - initiation, 326-327
- appareil de, 311-334
 - bactérien, 319-320
 - phage T4, 329-331
 - pré-initiation, intermédiaire de, 325-326
 - protéines, 322, 324
- ARN, à poliovirus, 332-334
- bidirectionnelle, 296-297
- brin, déplacement de, 332-334
- cercle roulant, 303-305
- chromatine active, 538-539
- chromatinienne, rupture de la structure, 557-558
- chromosome, 299-300
 - conservative :
 - nucléosome, 541
 - octamère d'histone, 539
 - cycle cellulaire, 299-300
 - définition, 97
 - déplacement ou boucle D, 305-307
 - discontinue, 317

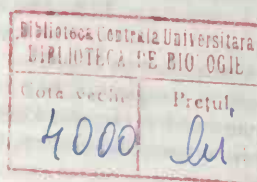
Réplication (*suite*)

- Ds, transposition, 609-611
- erreurs, 344
- fourche, 49, 51, 303, 305
 - brin principal, 317
 - définition, 296-297
 - réplisome, 312
 - structure, 539-540
- IS10, transposon, 595
- marquage radioactif, 297-298
- mutations, composants des, 322, 324
- réplicons linéaires, 331-334
- saltatoire, 438-439
- semi-conservative, 49
 - chromosome, 508-509
- semi-discontinue, 317
- séquentielle, 296-298
- structures, 303-307
- unidirectionnelle, 296-297
- virus, 615-616
- yeux, 296-297
- Réplisome, 312
- Réponse stringente, 240-241
 - catabolique, 236-240
- Répression :
 - définition, 220
 - multiple, 258
 - multivalent, 258
- Résolution :
 - réaction, 601
 - transposons, TnA, 603
- Résolvase, 601
- Restriction, carte de :
 - acides nucléiques, hybridation des, 78
 - ADN, fragments de clivage de l', 75-78
 - définition, 74
 - gènes en tandem, groupe de, 412-413
 - gènes interrompus, 385-387
 - kilopaires de bases (pb), 74
 - marqueurs génétiques, 78-83
 - polymorphisme, 78-81
 - paires de bases (pb), 74
 - techniques :
 - additivité complète, 76
 - digestion partielle, 77
 - double digestion, 75-78
 - marquage des extrémités, 77
- Réticulocytes, facteur d'initiation, 112
- Rétrocontrôle, inhibition par, 13
- Rétrorégulation, 449-450
- Rétrotransposons, 605

- Rétrovirus, 98
 activateur, 214
 éléments Ty de levure, 618-620
 fonction, 700
 oncogène, 701-703
 séquence cellulaire, transduction, 614-616
 transformant, 701-702
 transposition, analogues d'éléments de, 610-614
- Reversion :
 locus « white », 682-683
 mutations, 56
- Revertants, 35
- Rho, facteur, 249-250
E. coli, 259-261
- Rhyncosciara*, 626
- Ribonucléase :
E. coli, 447
 propriétés, 445-446
- Ribonucléase P, épissage de l'ARN, 465
- Ribonucléoprotéine, particule de, voir Ribosome
- Ribonucléoprotéine (RNP) :
 ARNm, traduction des, 164
 chromosomes en écouvillon, 514-515
- Ribose, 42
- Ribosome :
 activité GTPase — dépendante du, 110-111
 ARNm : localisation, 146-147
 traduction, 99-103
 assemblage :
 structure, 146-147
 topologie, 151-152
 assemblage *in vivo*, 150-152
 bactérien, structure, 146-147
 centres actifs, 154-156
 domaine de sortie, 154-155
 domaine de traduction, 154-155
 marquage d'affinité, 154-155
 cœurs, 150
 cytoplasmique, 145-146
 définition, 100
 détermination de la liaison au, 104
 facteur de stringence, 241
 liaison, 154
 modification, 154
 mutants sensibles au froid, 151
 mutation des composants, 152-154
 organelle, 145-146
 protéines, interaction avec les ARNr, 147-150
 reconstitution *in vitro*, 150-151
 ribonucléoprotéine, particules de,
- Ribosome (*suite*)
 145-147
sad, mutants, 151
 sites d'action, synthèse protéique, 105-106
 sous-unité, 106, 144-145
 synthèse protéique, sites de liaison, 106-109
 traduction, 144-157
 fidélité, 156-157
 vitesse de sédimentation, 106
- 30S :
 assemblage, 152-153
 complexe d'initiation, 106, 108-109
 reconstitution, 150-151
 structure, 145-146
- 50S, 108-109
 structure, 145-147
- 70S, 108
- Rifamycine, initiation de l'ARN, 187
- RI, particules, 150-151
- RI, plasmide, 676
- Rot, paramètre, 379
 gènes, nombre de, 380
- Rupture et réunion, mécanismes de, 27-29
- Rupture - fusion - pont, cycle de, 609
- S-adénosyl-homocystéine (SAH), 339-340
- S-adénosyl-méthionine (SAM), 339-340
- Salmonella* :
 diphasique, 586
hin, fonction, 588
 phase, variation de, 586-587
- ScARN, épissage, 479
- S. cerevisiae*: ADN chromosomique, séparation de l', 82-83
 ADN de levure, réarrangement de l', 430-431
 centromère, 512
 gènes interrompus, 388
 génome mitochondrial, 424-427
 grandes mitochondries, 427-428
 introns auto-épissable, 467
 réplicons, 303
 type sexuel, 634
 loci actifs et muets, 664-666
- Scrapie, 92
- scRNP, 479
- Scyrps, voir scRNP
- Secteurs jumeaux, analyse clonale, 610
- Segments D, 646-647
 réaction de jonction, 649
- Ségrégation :
 allèles, 19-20
 chromosome eucaryote, 509-512
 génome bactérien, 503-504
 marqueurs de restriction, 79-80
 post-méiotique, 576
 somatique, 423
- Ségrégation post-méiotique, 576
- Séquence intercalaire non transcrite, 412-413
- Séquences activatrices amont (UAS), 215
- Séquences exogènes, 632-634
- Sérum, dépendance du, 698
- Sexe, liaison du, 27
- Shine-Delgarno, séquence, 171-172
- "Shotgun", expérience, 361-362
- Signal, hypothèse du, 175-179
- Signal, particules de reconnaissance du, (SRP), 175-177
- Signal séquence, 175-179
- Simple-brin, réaction de passage au, 579
- SIR (régulateur d'information silencieux), 668-669
- Sites hypersensibles, 551-555
 activation des, 557-558
 loci MAT, 668-669
 stabilité, 552-553
- snARN :
 excision/épissage, 478-479
 fonction, 481
 U1, 479-480
 U2, 479
 U7, 488
- snRNP, 479
 ARN, réactions de maturation des, 481
 U, réaction *in vitro*, 488
 U7, 488
- Snurps, voir snRNP
- Solénoïde, 534-535
- Somatique :
 définition, 23
 mutation, 658-659
 recombinaison, 643-644
 ségrégation, 423
- Sonde, 359
- SOS, boîte, 348-349
 réponse, 348-350
- Souris :
 ADN transfecté, 634-635
 gène de l'amylase, 389-391
 gènes, marqueurs des, 72

- Sous-méthylation, 549-550
 Sous-réplication, 626
 Southern blotting, 360-361
 Spliceosome, 477-478
 Sporulation, 202, 204-205
 type sexuel de la levure, 665
S. purpuratus :
 ADN répétitifs, 373-374
 gènes, expression des, 383
 groupes des gènes histones, 413-414
 séquence de contrôle, 491
 Stéroïdes, transcription, 215-216
 Streptolydignes, initiation de l'ARN, 187
 Structure co-intégrée, 599-601
S. typhimurium :
 opéron tryptophane, 252
 phage SP6, 364
 suppresseurs, 140
 Sucre, molécules de, 4
 Supertours négatifs, 69-70
 Suppresseur, 51
 Suppression :
 définition, 35
 mutation petite, 431
 Suppression, intercistronique, voir Suppression
 Suppression d'un décalage du cadre ARNt, 142
 Synapse, 564
 interaction dépendant d'une, 684
 Synaptique, complexe, 564
 élément central, 564
 élément latéral, 564
 recombinaison, nœuds ou nodules de, 565
 Synthèse, 44-45
 Synthèse du brin conducteur, 327-328
 Synthétase, reconnaissance par la, 126-128
 Système de réparation des mammifères, 350
 Tautomériques, formes, 47
 Technique d'addition de queue poly (dA:dT), 357-358
 Télomère, 511-512
 gène, copie d'origine d'un, 672-673
 Tératomes, 676
 Terminaison :
 eucaryote, 251-252
 NusA, facteurs, 267
 palindromes, 249-250
 ribosome, blocage du, 253
 Terminaison (suite)
 synthèse protéique, 105
 codons pour, 120
 prématurée, 120
 transcription, catalyse par l'ARN polymérase, 185
 Termineuse :
 ARN polymérase, 248-251
 dépendant de rho, 249-251
 indépendant de rho, 249
 séquence, 185
 unités transcriptionnelles, 192-193
 Test croisé, technique du (Testcross), 22
 liaison (linkage), 27-28
Tetrahymena pigmentosa, gènes interrompus, 388
Tetrahymena pyriformis, sites hypersensibles, 554-555
Tetrahymena, séquences télomériques, 51
Tetrahymena thermophila :
 ARNr, organisation des, 461-463
 introns auto-épissables, 466
 Thalassémie, crossing-over inégal, 402-405
 α -Thalassémie :
 α -thal-1, 402
 α -thal-2, 402
 $\alpha\beta$ -Thalassémie, 405
 β -Thalassémie, groupe de gènes, 404-405
 Thréonine kinase, activité, 714-715
 Thymine, 41
 dimère, 342, 344
 système de réparation de E. Coli, 345
 Tissus polytènes, 626
 TMV, capsides, 501
 T_m :
 ADN renaturé, 375-377
 valeur, 59
 Topographie, 57-71
 Traduction :
 contrôle autogène, 242-244
 définition, 97
 Traduction arrêtée par hybridation, 359-360
 Transcriptase inverse, 611
 ARNm eucaryotique, 168-169
 Transcription :
 ADN, déroulement de l', 538
 élongation, nucléotides, 241
 ADN mitochondrial humain, 429-430
 ARN polymérase, catalyse par l', 184-185
 Transcription (suite)
 chromatine :
 active, 538-539
 domaines sensibles à l'ADNase, 545-546
 chromosome, structure du, 517-518
 contrôle, 491-493
 définition, 97
 élément *copia*, 621
 expression génique, 183-184
 facteurs :
 B, 212
 CTF, 212
 reconnaissance, séquence consensus de, 211-213
 sp1, 212
 TFIID, 212
 USF, 212
 facteurs eucaryotes, 213
 gène *ftz*, 687
 groupes ADNr, 416-417
 hnARN, contrôle de l', 489-491
 inhibition :
 enzymes cibles, 189
 promoteur de l'opéron ARNr, 241
 initiation :
 brin, séparation de, 200
 enhancers bidirectionnels, 213-216
 facteur sigma, 200-206
 facteurs, 213
 levure, type sexuel de la, 667-669
 nucléosome, organisation du, 542-543
 phage SPO1, 202-203
 phage lambda, 264-265
 phase végétative, 202
 protéine répresseur, 227-228
 réverse, 98
 sites hypersensibles, 552
 spécificité, 261-263
 transcrit primaire, 192-193
 pré-oncogènes, 714-715
 trp, opéron, 237-238
 ultrabithorax mutation, 691-692
 unité, voir Unité de transcription
 virus de la tumeur mammaire de la souris, 215
 Transcription inverse, 98
 Transcrit primaire, 192-193
 Transduction, rétrovirus, 614-616
 Transfection, 41
 Transférase, terminal, 357
 Transfert (blot), technique, 360-361
 ARNm, précurseurs nucléaires des, 472
 voir aussi techniques spécifiques

- Transformant, principe, 38-39
 Transformation, 711-715
 cellule, 698-699
 découverte, 38
 rétrovirus, 701-702
 virale, 699-701
 Transgéniques, espèces, 634
 Transition, mutation de, 52
 Translocation :
 définition, 27-29, 399-400
 leucémie myéloïde chronique, 709-710
 oncogène, 706-711
 synthèse protéique, 116-120
 liaison peptidique, 116-118
 Transplantation, antigène de, 660-662
 domaine cytoplasmique, 662
 domaine externe, 661-662
 région transmembranaire, 662
 Transport cotraduction, mécanisme de, 175
 Transposase, 592, 599-603
 Transposition :
 ARN-dépendant, 616-617
 conservative vs. répllicative, 597-598
 intermédiaires, 601
 élément de contrôle, 607-609
 élément Ds, répllication, 609-611
 éléments Ty, 619-620
 fréquence, 594
 immunité, 603
 intermédiaires, 599-603
 loci MAT, 669-670
 maïs, 605-607
 rétrovirus, cycle de vie du, 610-614
 supresseur-mutateur, 607-609
 Transposon :
 composé, 592-594
 modules IS10, 592-594
 cycle de répllication, 595
 Tn5, 596-597
 Tn10, 592-594
 module fonctionnel, 594-596
D. melanogaster, 620-622
 dysgénésie hybride, 622-624
 FB (repliement), 621-622
 mutations ponctuelles, 53
 séquences d'insertion, 590-592
 Transposon (suite)
 site *res*, 603
 structure cointégrée, famille TnA, 601
 Tn, 592-594
 TnA, famille, 601-602
 Tn3, 602-603
 Transvection, locus white, 683-684
 Transversion, 52
 Trypanosome, 664
Trypanosoma brucei, ADN
 réarrangement, 670-671
 Tryptophane, contrôle de l'opéron, 236-237
 Ty, éléments, 618-620
 Ty1, 618-619
 Tyrosine kinase, activité, 714-715
 Ty917, 618-619
 Ubiquitine, 547-548
 UH2A conjugués, 548
 Un gène : une chaîne polypeptidique, 32
 Unité cellulaire, 300
 Unité répétée, groupes de gènes, 412
 Unités extrachromosomiques, 269
 Unité transcriptionnelle :
 ADNr, organisation des nucléosomes de l', 543
 ARNr, répétition en tandem de l', 416-418
 bactérie, 160-161
 définition, 192-193, 412-413
 épissage, profil d', 474-475
 expression génique, 587
 point de départ aval et amont, 192-193
 régions distales et proximales, 192
 Uracile, 41
 Valeur C :
 définition, 367-368
 paradoxe, 367-369
 Van der Waals, définition des forces de, 9
 Variation de phase, 586-587
 Variégation, 607
 Vecteur de clonage pBR 322, 354-355
 Viroïde, définition, 91-93
 Virus :
 auxiliaire propriétés, 18
 (helper), 614
 réplication, 91-92
 réplication-défective, 614-616
 transformant, 614-616, 699-701
 viabilité, 39-40
 Virus MuSV, oncogènes, 702
 Virus polyome, brèches sensibles à la nucléase, 554
 Virus satellite de la nécrose du tabac (STNV), répllication, 91-92
 Virus SSV oncogènes, 702
 Virus SV40 :
 ADN super-enroulé, 531
 antigène grand T, 712
 complexes transcriptionnels, 543
 enhancer, 214-215
 épissage, profil d', 474-475
 minichromosomes, 554
 site de terminaison, 251
 trangénique, 635-636
 Virus transducteur, 614
 Virus VSV, 174
 Western blotting, 361
 White locus, 36, 682-685
 allèles, polymorphisme, 78-79
 mutation DZL, 684-685
 Xeroderma pigmentosum, 350
X. laevis :
 ARNm, 488
 ARN 5S, 216-218
 séquences, 419
 gènes, isolement des, 359
 globine, groupes des gènes, 406
 histone :
 gènes, organisation des, 414-415
 homeo box, 696-697
 réplicons, 303
 terminaison, 251
 traduction des ARNm, 166-167
 hypothèse du signal, 178-179
 Zygote, 19



BCU IASI / CENTRAL UNIVERSITY LIBRARY

Imprimé en Avril 1992
© Tardy Quercy S.A. - Bourges, imprimeur (n° 92-03-0089)
Flammarion et C^{ie}, éditeurs (n° 9989)
Dépôt légal : Mai 1992
Imprimé en France



FM 5041-92-IV

395,00 FF